Facial Animation
Driven by X-Ray Microbeam Data

November Scheidt and Howard J. Hamilton

Technical Report CS-2000-04
Apr., 2000

Department of Computer Science
University of Regina
Regina, Saskatchewan, CANADA
S4S 0A2

ISSN 0828-3494
ISBN 0-7731-0403-8

Facial Animation Driven by X-Ray Microbeam Data

November Scheidt and Howard J. Hamilton
Department of Computer Science, University of Regina
Regina, Saskatchewan, Canada S4S 0A2
E-mail: {nova,hamilton}@cs.uregina.ca

Abstract

In computer facial animation, consecutive frames are generated to create motion, or expression, on a computer modelled face. One facial animation technique, the performance-based animation technique, uses human data, such as a video recording that tracks the features of the face, to drive an animation. The present document describes a performance-based implementation called CASSI or Computer Animated Speech SImulator. CASSI uses human data in the form of 2D X-ray microbeam (XRMB) data to drive the animation of a 3D facial model. The 2D X-ray microbeam data contain coordinate values tracking the side view movement of eight gold pellets placed as follows: one on the upper lip, one on the lower lip, four on the tongue, and two on the jaw. The XRMB data track the movement of the pellets attached to human subjects who are performing speech-related tasks. The 3D facial model is an augmentation of the parameterized facial model developed by Parke.

CASSI was implemented as three versions: CASSI 1.0, CASSI 2.0, and CASSI 2.1. CASSI 1.0 was designed to integrate the XRMB data files with Parke's facial model. This integration included initializing the chin, palate, tongue, teeth, and lips of Parke's model, and animating the model, particularly, rotating the jaw, with the XRMB data. The emphasis of CASSI 2.0 was on lip movement, in particular, on rounding the lips using elliptical outlines, varying the thickness of the lips depending on the subject, and preventing the lips from running into the surface of the teeth using a parabolic track. In CASSI 2.1, instead of having the upper and lower lips of equal thickness, as in CASSI 2.0, the lip thickness for some subjects' was modified so that the upper lip is made smaller than the lower lip.

CASSI provides several research contributions. It animates a 3D face according to 2D X-ray microbeam data, handling jaw rotation and lip rounding. It models and animates a simple tongue, which most performance-based approaches do not include because the techniques used to record the movement of the face, for example, video camera, can not capture the movements of the tongue. CASSI also provides fairly good animations across several subjects without requiring manual adjustments to the parameters.

Chapter 1

Introduction

1.1 General Problem

Facial animation, in the most general sense, involves generating frames that create expression on an animated face. The facial animation may be of a person, or it may be of an animal or imaginary character. It may be based on the physical anatomy of the human face, or it may not have a physical basis. It may produce realistic results or caricatures. The common theme of facial animation is its focus on generating expression. This expression may be emotional expression, such as anger, fear, or happiness, it may be speech related expression, such as rounded lips for the "O" sound, or it may be general expressions, such as blinking or nodding.

Facial animation has been applied to several research areas. For the film industry, researchers have been focusing on producing realistically animated faces [26] [14] [19] [17], with the ultimate goal of producing virtual actors that recreate dead or existing actors, or create new characters. Other researchers [29] [6] are interested in using the animated synthetic face as an interface to allow people and computers to interact in a natural manner. An animated face can also be used in research concerning speech perception, including research on speech or hearing impairments; some applications include: as a supplement to the audio signal and for teaching lip reading to those with hearing impairments [9] [7] [16], and as a demonstration tool for those with speech impairments [16].

With all these applications, fast generation of the facial frames is advantageous. Two techniques are used to quickly generate frames: performance-based animation and speech synchronized animation. *Performance-based animation* uses human motions, which may be human facial movement, to drive the animation. *Speech synchronized animation* uses written text or auditory speech signals to drive the animation.

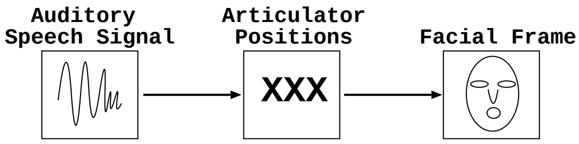


Figure 1.1: General Speech-Driven Approach

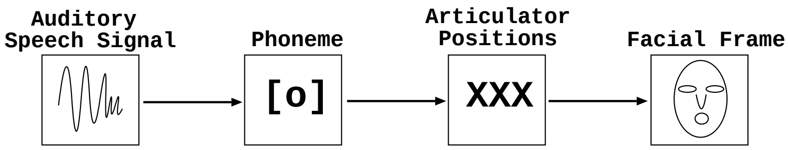


Figure 1.2: Phoneme-Based Approach

Our focus is on speech-synchronized animation. We would like to use a *speech-driven approach* where the auditory speech signal drives the animation. Figure 1.1 demonstrates this approach in its most general form. In the first step, the auditory speech signal is processed to find positions for the articulators. The *articulators* are the components of the mouth, namely, the tongue, lips, and jaw, that are responsible for producing speech. Using these articulator positions, the facial frame is then adjusted. Instead of this direct approach, some researchers [23] [18] use a *phoneme-based approach*, as shown in Figure 1.2. In this approach, a sample interval of the speech signal is processed to identify a phoneme or phoneme group in that interval. Using these phonemes or phoneme groups, the positions of the articulators are determined and the facial frame is accordingly adjusted. With this phoneme-based approach, interpolation is used to create a smooth motion between phoneme frames.

The general goal of our research is to contribute towards a system that does not depend on identifying phonemes and is closer to the approach shown in Figure 1.1. There are two advantages of this approach. First, the intermediate step of identifying phonemes is removed; this provides an advantage since clear samples of phonemes are often difficult to find in fast speech. Second, it eliminates the need for interpolation since "in between" phoneme frames will also be accounted for as unique placements of the articulators.

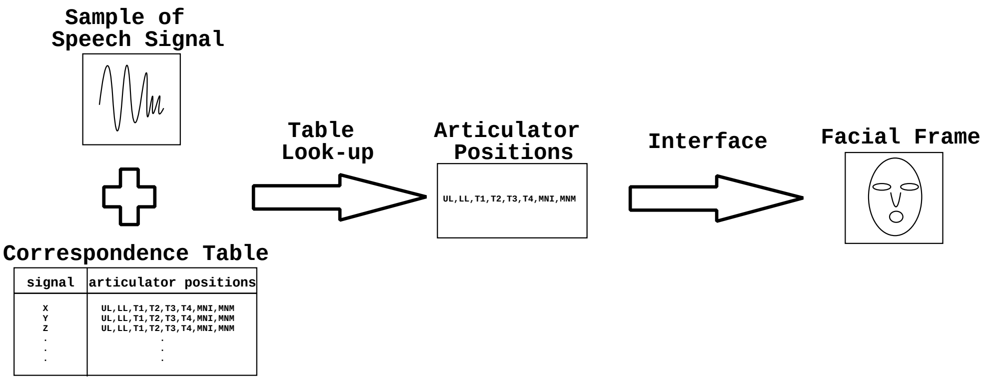


Figure 1.3: Correspondence Table Approach

Figure 1.3 shows our proposed *correspondence table approach*. In this approach a correspondence table is used to associate the speech sample with the placement of the articulators. Given a sample interval of the entire speech signal, the correspondence table is accessed to find the closest matching sample signal. The articulator positions associated with this closest match are then output. Next, an interface is used to map these articulator positions to one facial frame.

To implement the correspondence table approach, we require data that include both articulator positions and corresponding speech signal recordings. The X-ray microbeam data [30] provide these two corresponding parts. The speech signal has been recorded as speakers perform tasks such as reading sentences, words, or paragraphs. The positions of the lips, jaw, and tongue have also been recorded. The movement of these articulators is captured by an X-ray microbeam that tracks the movement of eight gold pellets placed as follows: one on the upper lip, one on the lower lip, two on the jaw, and four along the centerline of the tongue. These pellets are tracked from the side view and, thus, have 2D (x and y) coordinates associated with given points in time.

The correspondence table approach takes an auditory signal and a correspondence table and produces articulator positions. These articulator positions, in the form of pellet coordinates, are used to drive the animation. Two smaller components are used in this approach. Figure 1.4 shows the learning component. This component accumulates a table of corresponding speech signals and articulator positions. The resulting output is a correspondence table, which describes a functional mapping from each given interval of speech to the corresponding positions of the lips, tongue, and jaw. The second component, shown in Figure 1.5, is an interface where the articulator positions drive the facial animation.

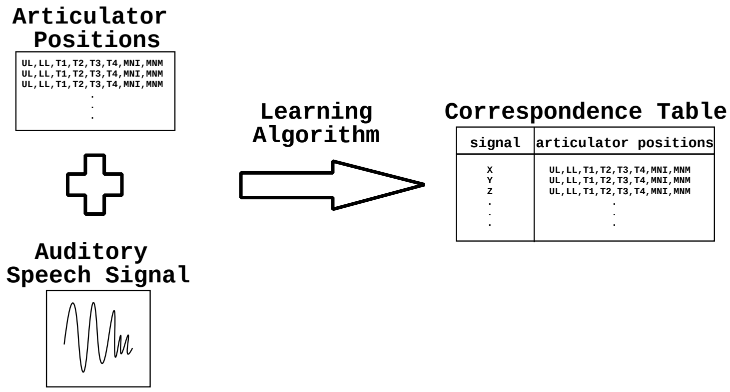


Figure 1.4: Learning Component of Our Approach

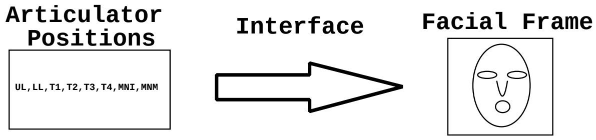


Figure 1.5: Animation Interface Component of Our Approach

1.2 Specific Problem

The scope of this research is limited to the component shown in Figure 1.5. Articulator positions are obtained from X-ray microbeam data. Designing this component requires the creation of an interface between the 2D X-Ray microbeam pellet coordinates and the 3D facial model. This component does not require synchronizing the auditory speech to the animated face. It also does not require reflexes, such as blinking, or include underlying expressions, such as anger or happiness.

The CASSI (Computer Animated Speech SImulator) system was developed to provide an interface between the X-ray microbeam data and the animated face. CASSI uses Parke's facial model [23] for its underlying topology and parameters. CASSI was developed as a series of three augmenting versions. The first version provided the following: initialization of the face, which included drawing a palate, adjusting the jaw, and setting the tongue, lips, and teeth in an initial position according to the X-ray microbeam data; animation of the lips, teeth, jaw and tongue based on the changes in the pellet coordinates from the initialized location; and rotation of the jaw and surrounding vertices based on the movement of two jaw pellets. The second version emphasized lip movement, in particular: rounding the lips, preventing the lips from running into the teeth, and creating an individual lip thickness for each speaker. The third version incorporated an improved technique for setting upper and lower lip thicknesses for each speaker.

Three original contributions are made by this research. First, CASSI animates the movement of the tongue according to real human motion. To our knowledge, other performance-based animations do not include tongue animation. Second, it provides animation across several subjects without requiring manual adjustments in the parameter values. Other performance based approaches seem to be limited to only one subject, or their descriptions imply that some manual work is required to map the subjects' movements to a facial animation. Third, CASSI translates 2D data obtained from a side view of a human face into the movement of an animated 3D face. Other performance-based approaches use a frontal view, or the combined input of a frontal and a side view, to drive their animations.

1.3 Overview of Chapters

The remainder of this document is organized as follows. In Chapter 2, we provide a survey of facial animation techniques to illustrate their most significant characteristics and the reasons we have chosen our facial animation approach. In this survey, we discuss the following techniques: image-based key frame animation, parametric key frame animation, performance-based animation, and speech synchronized animation.

Our approach involves two components, Parke's model and the X-ray microbeam data. Parke's 3D coordinate system is discussed in Section 3.1 with a focus on the topology, parameters, and jaw rotation procedures inherited by CASSI. The X-ray microbeam system is discussed in Section 3.2 to explain the method of data collection, the coordinate system, which was developed for several subjects, and the resulting data.

Chapter 4 discusses the CASSI software system as it was implemented as a series of three versions. CASSI 1.0, which focused on initialization, animation, and jaw rotation of Parke's original model, is discussed in detail. The underlying mathematical basis and assumptions made in CASSI 2.0 are provided for a better understanding of the revised lip motion created in this version. The new lip motion in CASSI 2.0 includes: rounding the lip, preventing the lips from running through the teeth, and creating a unique lip thickness for each subject. CASSI 2.1 is discussed with regard to modifications made to the lip thickness to provide a better view of the mouth.

In Chapter 5, a subjective evaluation is presented of the placement and movement of the teeth, tongue and lips in CASSI 1.0. The lip movements in the three versions of CASSI are compared with regard to the improvements made in overall lip motion, to the lip rounding created, and to the movement of the lips during large jaw rotation.

Chapter 6 summarizes the contents of this document, identifies the original work contributed by this research, and discusses future work.

Chapter 2

Facial Animation Techniques

In general, animation is created by displaying a succession of images, or frames, which differ slightly from each other. For facial animation, at least four major techniques are used to create the moving images of the face. These techniques include image-based key frame animation, parametric key frame animation, performance-based animation, and speech-synchronized animation. The first two techniques, image-based key frame animation and parametric key frame animation, both depend on *key frames*, which specify the desired facial expression for certain points in time, and *interpolation*, which generates the "in-between" frames. The goal of the other two techniques, performance-based animation and speech-synchronized animation, is to quickly generate frames by using, respectively, human action and speech to drive the animation.

Image-based key frame animation and parametric key frame animation rely on the concepts of key frames and interpolation. Several key frames (or *key poses*) are chosen for the animation and then interpolation is applied to generate a smooth motion between key frames. For instance, one key frame may be of a woman with her mouth open, and a second key frame may be of the woman with her mouth closed. The interpolation algorithm is then applied to generate frames which are somewhere between open and closed. Image-based key frame animation interpolates between entire key frame images. Parametric key frame animation interpolates parameter values.

To quickly generate frames, performance-based animation and speech-synchronized animation are used. Performance-based animation uses human action to drive the model. Approaches to performance-based animation include: expression mapping, which involves matching real human expressions to a data base of character expressions; puppetry, which involves moving an animated computer character through hand motions such as a puppet master would use to manipulate a puppet; and tracking the face over time, which may involve tracking markers placed on the face and moving the animation to correspond to these markers. In speech-synchronized animation, the frames are quickly generated based on human speech. Two major approaches to speech-synchronized animation include text-driven and speech-driven. Generally, in these approaches, the phonemes (or individual speech sounds) are determined, the key poses resulting from these phonemes are used in the facial animation, and an interpolation algorithm is used to create a smooth transition between phonemes.

Section 2.1 briefly describes image-based key frame animation and provides examples of short animations that have used this technique. Section 2.2 describes parametric key frame animation and three approaches used to create the parameters: direct parameterized models, pseudo muscle-based models, and muscle-based models. In Section 2.3, performance-based animation is discussed with emphasis on techniques used to track the face over time. Section 2.4 discusses text-driven and speech-driven speech-synchronized animation. The final two sections summarize these four animation techniques and provide a brief description of the technique used in this research.

2.1 Image-Based Key Frame Animation

For image-based key frame animation, an image, corresponding to a key frame, is specified (digitized) at a certain point in time and then another image is specified several frames later. A computer algorithm is then used to interpolate between these digitized images. Image-based key frame animation is used in the films *Sextone for President* [25] and *Don't Touch Me* [10]. Both films contain facial animation created by digitizing the face in several expressions and then interpolating between these expressions. According to Magnenat-Thalmann and Thalmann [20], this technique requires sufficient key frames to obtain realistic results. In addition, according to Parke and Waters [23], image-based key frame animation is labor intensive since the geometry of the face must be calculated for each key frame.

2.2 Parametric Key Frame Animation

Parametric key frame animation involves parameterized facial models. Any object can be described by a set of parameters, which determine its shape and properties. For instance, a cube can be described by parameters for its length, color, and material. Similarly, a face can be specified by a set of parameter values for its height, width, length, distance between eyes, etc. Because the entire facial geometry does not need to be specified, parameterized facial models provide data compression, and a relatively convenient way of specifying desired changes in the face.

In parametric key frame animation, the animator creates key frames by specifying the appropriate set of parameter values at specific points in time. The parameter values can then be interpolated, producing the in-between frames. Parametric key frame animation has an advantage over image-based key frame animation because only the parameters are interpolated rather than the entire image. However, according to Magnenat-Thalmann and Thalmann [20], a good choice of parameters is essential to guarantee a good animation.

Three approaches to parametric key frame animation include: direct parameterized models, pseudomuscle-based models, and muscle-based models. Each approach has a unique system of parameters. For instance, direct parameterized models use parameters for scaling, translation, interpolation, and rotation. Pseudomuscle-based models use a few control parameters that correspond approximately to a muscle. Lastly, muscle-based models use parameters to emulate muscles.

2.2.1 Direct Parameterized (Topological) Models

In direct parameterized models, parameters are used for rotation, scaling, translation, and interpolation of specified regions of the face. These models have little theoretical basis and do not pay careful attention to facial anatomy. An example of a direct parameterized model is Parke's model.

Parke's model [22] consists of about 400 vertices which are connected in a polygonal network, or topology, of about 300 polygons. Figure 2.1 shows Parke's model; the right side shows the polygon topology, and the left side shows the face with color and shading (shown in monochrome). The topology, or connections between vertices, remains constant, but the position of each polygon vertex varies according to values of the parameters. As the vertex positions change, the polygonal surfaces flex and stretch causing the face itself to change. Thus, by adjusting the parameters, a single topology can display many expressions and a wide range of individual faces.

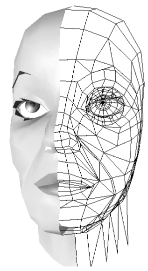


Figure 2.1: Parke's Model Showing Polygon Topology.

Table 2.1 shows a few of the 50 parameters used in Parke's model [23]. This table separates the parameters into four categories: mouth region parameters, interpolation parameters, scaling parameters, and translation parameters. In the mouth region, the jaw rotation parameter has a starting value of 0, which yields the face shown in Figure 2.2(a). By setting the jaw rotation to 20, as demonstrated in Figure 2.2(b), points in the

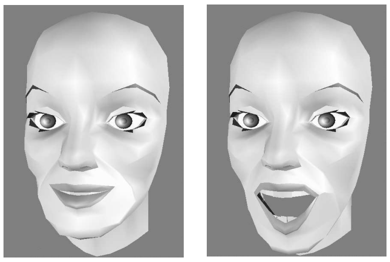


Figure 2.2: Parke's Model with the Jaw Parameter (JP) Set to 0 and 20.

lower lip, chin, neck and cheek rotate around a fixed point (representing the joint of the jaw). For the second category, interpolation, an example is the eye opening parameter. In this case, two extreme sets of points are given, one set of points for the open eyelid, and another for the closed eyelid. For a wide open eyelid, the eye opening parameter value is set to 1. For a closed eyelid, the eye opening parameter value is set to 0. Values between 0 and 1 specify interpolations between open and closed. The third category, scaling, controls the relative size of facial features. An example of scaling is the Head Y Scale. Setting the Head Y Scale to 2.0 (twice the amount of the starting value), creates a face that is twice as wide as the original. Translation, the fourth category, results in moving a collection of vertices as a group. For instance, setting Nose Tip Z Offset to 30 moves points in the tip of the nose closer to the eyes.

Category Number	Parameter	Parameter Label	Starting Value Where Applicable	Value Range
Mouth Region	4	Jaw Rotation	0.0	$0.0 \leq value \leq 20.0$
	12	Mouth Y Scale	1.0	
	13	Mouth Interpolation	1.0	$0.0 \leq value \leq 1.0$
	16	Mouth Corner X Offset	0.0	
	17	Mouth Corner Y Offset	0.0	
	18	Mouth Corner Z Offset	0.0	
	21	Lower Lip "f" Tuck	0.0	
Inter-polation	1	Eye Opening	0.9	$0.0 \leq value \leq 1.0$
	2	Eyebrow Arch	0.5	$0.0 \leq value \leq 1.0$
Scaling	7	Head X Scale	1.0	
	8	Head Y Scale	1.0	
	9	Head Z Scale	1.0	
	10	Nose Tip Y Scale	1.0	
	11	Nose Bridge Y Scale	1.0	
	35	Chin-to-Mouth	1.0	
	36	Chin-to-Eye	1.0	
	37	Eye-to-Forehead	1.0	
Trans-lation	26	Nose Tip X Offset	0.0	
	27	Nose Tip Z Offset	0.0	
	31	Chin X Offset	0.0	
	32	Chin Z Offset	0.0	

Table 2.1: Some Parameters in Parke's Original Model [23]

2.2.2 Pseudomuscle-Based Model

In real human faces, a complex interaction of muscles, bones and facial tissue is involved in producing facial expressions. The pseudomuscle-based model is not concerned with these complex interactions or with the details of facial anatomy; instead, it is concerned with emulating the movement of some of the basic facial muscles. Thus, the parameters in pseudomuscle-based models are designed to approximate the movement of the facial muscles.

Instead of simulating the movement of the facial tissue in relation to contracting muscles, pseudomuscle-based models use "geometric deformation operators". These geometric deformation operators control or change the shape of the face. For instance, one example of a pseudomuscle-based model involves freeform deformation [23]. Freeform deformation can be described through a physical analogy. Imagine a flexible, three-dimensional object that has been embedded in a block of flexible plastic. By stretching, bending, or twisting the block of flexible plastic, the object inside is also deformed. The change in the embedded object is a natural result of the change in the entire block. Thus, in pseudomuscle-based models, movement is created by defining a volume around the face or a section of the face. When this "control" volume changes, the points included in the volume also change.

Figure 2.3: Marilyn Monroe from *Rendez-vous à Montréal* [19]

Magnenat-Thalmann et al. [19] developed a pseudomuscle-based model, which they used in their film *Rendez-vous à Montréal* for the synthetic actors Marilyn Monroe (shown in Figure 2.3) and Humphrey Bogart. Their pseudomuscle-based model uses Abstract Muscle Action procedures (AMA procedures). Each AMA procedure is responsible for a facial parameter or parameters corresponding approximately to a muscle.

Number	AMA procedure	Range for corresponding values
1	VERTICAL_JAW	$0 \leq value \leq 1$
2	CLOSE_UPPER_LIP	$0 \leq value \leq 1$
3	CLOSE_LOWER_LIP	$0 \leq value \leq 1$
4	COMPRESSED_LIP	$0 \leq value \leq 1$
6	MOUTH_BEAK	$0 \leq value \leq 1$
7	RIGHT_EYELID	$-1 \leq value \leq 1$
8	LEFT_EYELID	$-1 \leq value \leq 1$
9	LEFT_LIP_RAISER	$0 \leq value \leq 1$
10	RIGHT_LIP_RAISER	$0 \leq value \leq 1$
11	LEFT_ZYGOMATIC	$0 \leq value \leq 1$
12	RIGHT_ZYGOMATIC	$0 \leq value \leq 1$
23	MOVE_RIGHT_EYE_HORIZONTAL	$-1 \leq value \leq 1$
24	MOVE_RIGHT_EYE_VERTICAL	$-1 \leq value \leq 1$
25	MOVE_LEFT_EYE_HORIZONTAL	$-1 \leq value \leq 1$
26	MOVE_LEFT_EYE_VERTICAL	$-1 \leq value \leq 1$
27	RIGHT_RISORIUS	$0 \leq value \leq 1$
28	LEFT_RISORIUS	$0 \leq value \leq 1$
29	MOVE_RIGHT_EYEBROW	$-1 \leq value \leq 1$
30	MOVE_LEFT_EYEBROW	$-1 \leq value \leq 1$

Table 2.2: Important Abstract Muscle Action Procedures [19]

Table 2.2 shows a few of the AMA procedures. The CLOSE_UPPER_LIP and CLOSE_LOWER_LIP procedures listed in the table close the lips when they are open. Each lip may move independently from the other. These procedures move the lips towards each other to the best location for contact, which is defined by the height of the corners of the lips. The shape of each lip is approximated by a curve defined by three points: the two corners of the lips (LEFTVERT and RIGHTVERT) and one point on the center of the lip (CENTERVERT). Two separate CENTERVERT's define the curve of the upper lip and the lower lip. As a second example, the LEFT_ZYGOMATIC and RIGHT_ZYGOMATIC procedures listed in the table simulate the movement of the zygomatic muscles, which are responsible for smiling. These two procedures react on facial vertices within an action volume (generally defined by a box). An initial vertex in the action volume is translated by some vector (DISP), and all the other vertices inside of the same volume are translated by some fraction of DISP depending on their location within the action volume.

2.2.3 Muscle-Based Model

Muscle-based models are designed to closely mirror human anatomy by including representations of layers of skin, bone, and muscle. The parameters in these models relate to actual muscles. The idea is to manipulate facial expression by contracting/relaxing a simulated muscle through changes in its parameter values. Two muscle-based models are discussed: Water's muscle model, which consists of muscle vectors with zones of influence; and the physics-based model, which represents the skin in detail, including its underlying layers.

Waters' Muscle Model

Waters [28] integrated muscles as parameters into his facial model so that he could generate such expressions as anger, fear, surprise, disgust, happiness, and sadness. He used the Facial Action Coding System (FACS) as a guideline for determining which muscles needed to be created and how much "muscle action" was required to generate an appropriate expression. Each muscle was modelled as a vector with a zone of influence.

The Facial Action Coding System (FACS) [11] was developed by Ekman and Friesen, psychologists of non-verbal communication, as a means of identifying facial movement independent of the speaker. Unlike previous systems, FACS was based on the movement of the muscles. To develop the system, Ekman and Friesen first learned how to separately fire the muscles in their own faces. They then examined photographs of the individually fired muscles to determine which muscles resulted in unique appearance changes. The idea was to find a minimal set of *action units* (AU). Each action unit, as shown in Table 2.3, generally corresponds to the action of one muscle. However, sometimes, the appearance changes resulting from the firing of two or three individual muscles were similar; in these cases, muscles were grouped together in one action unit. An example of one action unit involving a group of muscles is AU 39 (Nostril Compressor), which involves two muscles around the nose. In addition to one action unit involving a group of muscles, one muscle may have more than one action unit. For instance, the *frontalis* muscle, which raises the brow, was separated into two action units (AUs): AU 1 (the inner brow raiser) and AU 2 (the outer brow raiser).

Certain groups of Action Units act together to form the basic facial expressions (anger, fear, surprise, disgust, happiness, and sadness), as described by Ekman [12]. For instance, Waters [23] used Action Units 6, 12, and 11 to compress the cheeks, raise the corners of the lips, and widen the nostrils, respectively. The resulting face emulated the expression of happiness. To emulate emotions, Waters first created muscles in his model corresponding to the Action Units. The results of the real Action Units are used as guidelines for the parameter values used in his muscle model.

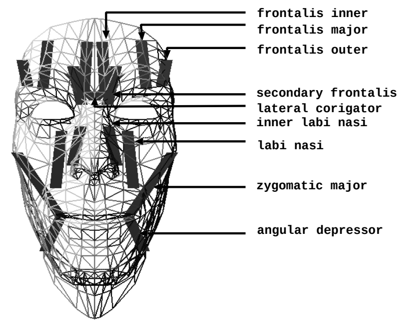


Figure 2.4: Waters' Model Showing Muscles

Waters based the muscles in his model on the functioning of real human muscles. For most human muscles, one end has a point of bony attachment that remains static, and the other end is embedded in soft tissue of the skin that contracts when the muscle is operated. Waters emulated the real muscles using muscle vectors with one end as the static point of attachment and the other end as the point of insertion into the skin. To model the visco-elastic nature of the skin, each muscle vector had a zone of influence such that the movement of the muscle vector affected the position of those neighboring nodes within the zone of influence.

In the code provided by Waters [23], the following muscle vectors are included: left and right zygomatic major, left and right angular depressor, left and right frontalis inner, left and right frontalis major, left and right frontalis outer, left and right labi nasi, left and right inner labi nasi, left and right lateral corrugator, and left and right secondary frontalis. These muscles are shown in Figure 2.4 as thick lines. Figure 2.5 shows that the effect of constricting the zygomatic major muscles is a smile.

AU number	FACS name	Muscular basis
1	Inner brow raiser	*Frontalis, pars medialis*
2	Outer brow raiser	*Frontalis, pars lateralis*
4	Brow lowerer	*Depressor glabellae; depressor supercilii;* corrugator
5	Upper lid raiser	*Levator palpebrae superioris*
6	Cheek raiser	*Orbicularis oculi, pars orbitalis*
7	Lid tightener	*Orbicularis oculi, pars palpebralis*
8	Lips toward each other	*Orbicularis oris*
9	Nose wrinkler	*Levator labii superioris, alaeque nasi*
10	Upper lip raiser	*Levator labii superioris, caput infraorbitalis*
11	Nasolabial furrow deepener	*Zygomatic minor*
12	Lip corner puller	*Zygomatic major*
13	Cheek puffer	*Caninus*
14	Dimpler	*Buccinator*
15	Lip corner depressor	*Triangularis*
16	Lower lip depressor	*Depressor labii inferioris*
17	Chin raiser	*Mentalis*
18	Lip puckerer	*Incisivii labii superioris; incisivus labii inferioris*
20	Lip stretcher	*Risorious*
22	Lip funneler	*Orbicularis oris*
23	Lip tightener	*Orbicularis oris*
24	Lip pressor	*Orbicularis oris*
25	Lips part	*Depressor labii,* or relaxation of *mentalis* or *orbicularis oris*
26	Jaw drops	*Masseter;* temporal and internal *pterygoid* relaxed
27	Mouth stretches	*Pterygoids;* digastric
28	Lips suck	*Orbicularis oris*
38	Nostril dilator	*Nasalis, pars alaris*
39	Nostril compressor	*Nasalis, pars transversa and depressor septi alae nasi*
41	Lids droop	Relaxation of *levator palpebrae superioris*
42	Eyes slit	*Orbicularis oculi*
43	Eyes close	Relaxation of *levator palpebrae superioris*
44	Squint	*Orbicularis oculi, pars palpebralis*
45	Blink	Relaxation of *levator palpebrae* and contraction of *orbicularis oculi,* pars palpebralis
46	Wink	*Orbicularis oculi*

Table 2.3: Single Action Units [11]

Figure 2.5: Waters' Model with Relaxed and Constricted Zygomatic Muscles

Physics-Based Facial Model

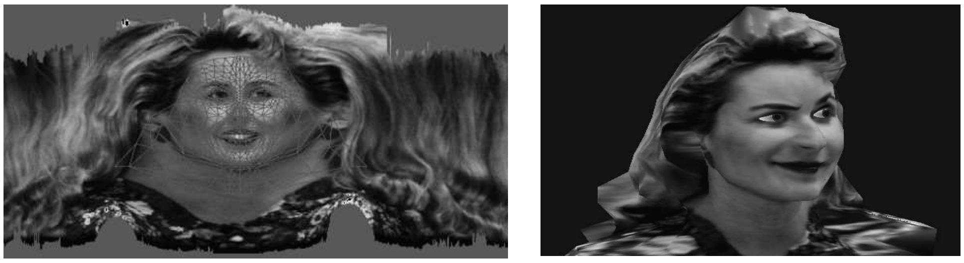
Physics-based facial models are a more complex muscle model than Waters'. Lee et al.[17], with their physics-based facial model, added several things to make the model more life-like. They created a new method of constructing the facial mesh that involved adjusting a generic mesh to fit the data of a real face. Also, they developed a physics-based facial mesh that was meant to model the various layers of skin with springs between and among layers. In addition, the skin was designed to preserve its volume and was constrained to slide over the skull structure.

The new method of constructing the facial mesh uses data from a laser scanner. The laser scanner circles around a person's head to acquire range and reflectance information. Each range value represents the distance from the scanner to the head. The reflectance data are RGB values. Figure 2.6(a) shows, in 3D, the range data for a person called "Heidi", and Figure 2.6(b) shows the same person's RGB values as a monochrome image. To find the "edges" of the face, a modified Laplacian operator is applied to the range data. The generic mesh is then fit according to the range image and its Laplacian field function. Figure 2.7(a) shows the adapted mesh superimposed on the RGB data. Finally, Figure 2.7(b) shows the resulting 3D face with a quizzical expression.

Figure 2.6: "Heidi" Data from Scanner.

The physics-based facial mesh models the layers of skin with springs between and among layers. The human skull is covered by five distinct layers: epidermis, dermis, subcutaneous connective tissue, fascia, and muscles. The physics-based model is designed in accordance with the skin's layered structure. Figure 2.8 illustrates the several layers of this new tissue model. Nine nodes represent three surfaces; nodes 1, 2, and 3 represent the epidermal surface, nodes 4, 5, and 6 represent the fascia surface, and nodes 7, 8, and 9 represent the skull surface. Between the three surfaces, lie two layers: the dermal fatty layer and the muscle layer. Springs connect the nodes together: epidermal springs connect nodes 1, 2, and 3; fascia springs connect nodes 4, 5, and 6; dermal-fatty layer springs connect epidermal nodes to fascia nodes; and muscle layer springs connect fascia nodes to skull surface nodes. The simulated muscles are embedded into the muscle layer of the skin with a fixed point at the skull surface and attachments at the fascia nodes as they run through several tissue elements.

To ensure that the skin moves in a realistic manner, two constraints have been placed on the model. The



(a)Adapted Mesh Superimposed on the Texture Map **(b)Resulting Heidi with a Quizzical Expression**

Figure 2.7: Heidi Face.

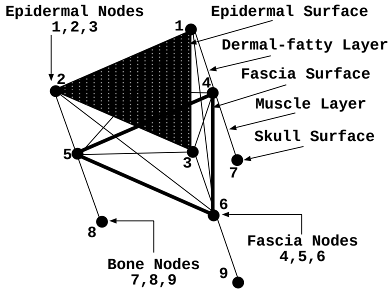


Figure 2.8: A Single Facial Tissue Element [17]

first constraint is a volume constraint, which assumes that the facial tissue is primarily water-based and, therefore, preserves its volume under deformation. This constraint causes the skin to bulge near the end of the muscles and depress near the central stretching area. The second constraint is a skull penetration constraint. This constraint ensures that the facial muscles slide over the skull and that the fascia nodes do not penetrate the skull.

2.2.4 Summary of Major Parameterized Facial Models

Parametric key-frame animation uses parameters to define desired key frame images and then uses interpolation to determine the in-between frames. Three approaches to parametric key-frame animation were discussed in the previous sections: direct parameterized, pseudomuscle-based, and muscle-based. For the direct parameterized approach, the parameters are based on scaling, translation, and interpolation; for pseudomuscle based, the parameters correspond approximately to muscles; and for muscle based, the parameters emulate real human muscles. Table 2.4 summarizes these three approaches and lists some of their advantages and the disadvantages. The first column identifies the approach, the second gives a brief description of the approach, the third provides an example of a model that uses this approach, the fourth gives its advantages, and the fifth gives its disadvantages.

Muscle-based models are not used for real-time applications because of the time required to generate each frame of the animation. This approach has, however, been used in several animated films, such as "Tin Toy"[14], "Toy Story", "A Bug's Life", and "Antz"[26][3]. For real-time applications, approaches such as Parke's seem to be preferred because of the low computational complexity.

2.3 Performance-Based Animation

Creating facial animation using parametric key frame animation is a tedious process. The facial shape at key frames must first be determined and then the proper combination of parameter values must be chosen to match the desired facial shape. Once the key frames have been created, the proper timing must be coordinated so that the face moves at the proper time.

Approach	Description	Example	Advantages	Disadvantages
Direct Parameterized	– parameters are based on scaling, translation, interpolation, and rotation	Parke's	– simple to use – easily accessible – easy to adjust parameters to create individual shaped faces and different speech movements – low computational complexity	– model appears plastic-like in the sense that the "skin" surface does not stretch and compress
Pseudomuscle-based	– emulates muscle action using geometric deformation operators	Magnenat-Thalmann's	– less complicated than the muscle-based model	– may have difficulty making individual faces. For Magenant-Thalmann's model, each character geometry was taken from a plaster cast, and then extreme parameter values were chosen so that each character would have a unique movement
Muscle-based	– the actions of the muscles are emulated with vectors – parameters are based on FACS (Facial Action Coding System)	Waters'	– realistic results. The skin stretches and compresses with muscle movement. Texture mapping can add to the realism of skin texture	– more complex than other methods – more difficult to achieve desired mouth shapes using muscle actions

Table 2.4: Summary of the Major Parameterized Facial Models

The goal of performance animation is to quickly generate the animation frames by using the motion of live performers or puppeteers to drive the animation. Human action is captured and mapped onto a computer animated facial model. Although there are other approaches, including expression mapping [15] [4] and puppetry [23], the following sections specifically discuss tracking facial features using automatic spot tracking, snakes, and electromyographic recordings. Each of these techniques tracks the facial movement of a human actor and uses this movement to drive the computer animation.

2.3.1 Automatic Spot Tracking

Williams [31] created a performance-driven facial animation system in which he videotaped and tracked the movement of markers placed on an actor's face. The movement of these markers was used to drive a facial model that had surface deformations similar to pseudo-muscle based models. Regions to be deformed on the 3D model were specified based on the location of the muscles and the location of the markers placed on the performance actor. Global head movements were ignored so that the problem was one of tracking x and y coordinates for the markers.

To track the markers on the face, several steps are taken. First, reflective markers are stuck to the face of the performance actor such that the markers never touch and are never obscured by another part of the face. The camera and the lighting are placed so that only bright spots are seen moving on a dark background. Then a video recording is taken while the actor produces facial expressions. To start tracking, the spots are manually selected in the first video frame. Next, a window (slightly larger than the spots) centered around the manually selected spot is used to compute the *center of gravity*, an estimate of the center of light intensity in the window. The window is then re-centered according to this computed center, and further refined through repeated computations of the center of gravity. After several iterations, the center of gravity represents the location of the marker. For each frame of the recording, the center of gravity for each marker is found. The displacements of these markers over time are mapped onto the facial model by warping specified areas of the face.

2.3.2 Snakes

Snakes (or *active contours*) track feature lines and boundaries in an image. They are two-dimensional curves or splines that are placed on or close to specific features, such as the eyebrows or lip boundaries. These snakes follow the movement of the features using 2D potential functions, which are obtained through image processing performed on the digitized image frames. The 2D potential functions have *ravines* (extended local minimum), which correspond to intensity changes associated with facial features such as the eyebrows,

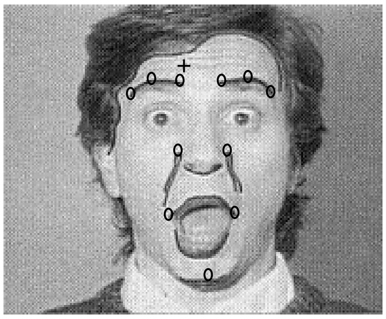


Figure 2.9: Snakes and Fiducial Points for a Surprised Expression

mouth, and chin. The snakes "slide downhill" and come to an equilibrium at the bottom of the nearest ravine.

Terzapoulos and Waters [27] used snakes to estimate the muscle contractions from a video sequence of an expressive human face. Using the estimated muscle contractions, they were able to control the muscles of the physics-based facial model described in Section 2.2.3.

Terzapoulos and Waters' implementation involved several steps. Before recording the actor, human features such as lips, eyebrows and nasolabial furrows were enhanced using make-up. Then, the actor was videotaped performing facial expressions in frontal view before the camera. Image processing was then applied to the sequence of images to create broadened ravines, which would attract the snakes from a distance. For the first frame of the digitized sequence, the nine snakes (shown as dark lines superimposed on the image in Figure 2.9) were initialized using the mouse. These snakes were place on: the hairline, left and right eyebrows, tip of the nose, chin boss, upper and lower lips, and the nasolabial furrows. Once the snakes locked into the ravines, a *head reference frame*, and 11 major points along the snakes (shown as open circles in Figure 2.9) were used as guidelines for adjusting the muscles. The head reference frame (shown as a cross bar above the speaker's right eyebrow in Figure 2.9) was used for removing global head movements and was determined from the average position of the hairline contour. The 11 major points included: three points along the eyebrow contour, which determined the contractions of the left and right inner, major, and outer frontalis; two endpoints of the upper lip contour, which determined the contractions of the left and right zygomatic major and the angular depressor; the upper-most positions of each of the nasolabial furrows, which determined the contractions of the inner labi nasi; and the average positions of the chin boss contour, which determined the jaw rotation. Once all the contractions were adjusted via the muscle parameters in the physics-based facial model, the next frame in the sequence was introduced, the snakes were allowed to slide downhill to conform to the shape of the ravines, equilibrium was attained, and the muscle contractions were again determined. This process of introducing frames, allowing the snakes to slide downhill, and determining muscle contractions, or parameter values, was repeated for the entire video sequence. These frame by frame adjustments to the parameter values created the animation in the physics-based facial model.

2.3.3 Electromyographic (EMG) Recordings

To study speech perception and production, Munhall [21] uses processed electromyographic (EMG) recordings from a subject's facial muscles to drive an extension of the physics-based muscle model mentioned in Section 2.2.3. The electromyographic recordings are obtained through the signals of needle electrodes inserted into the skin near specific muscles. These signals indicate when the muscle fiber or nerve is simulated and can be mapped into the facial model's muscle parameters. The disadvantages of this method are that a doctor or trained individual must insert the needles, the needles are a painful and intrusive method of tracking, and the resulting electromyographic recording may include noise. The advantage of this method is that it captures the contractions of the muscles themselves instead of capturing the surface characteristics of the face.

2.4 Speech-Synchronized Animation

In *speech-synchronized animation*, the frames of the animation, are directly connected to, or synchronized with, samples of speech. Changes in the speech samples, result in automatic changes in the animation. The two major approaches to producing speech-synchronized animation are text-driven and speech-driven. In *text-driven animation*, an arbitrary plain ASCII text is input, and synthetic speech and a corresponding animated face are output. In this approach, speech and images are created simultaneously. For *speech-driven animation*, a recorded speech sample is analyzed to determine the speech segments, and using these speech segments, a corresponding animation is produced. In this approach, the images are created based on an existing speech track.

2.4.1 Text-Driven Systems

In the text-driven approach, text is entered into the system and synthesized speech and a synchronized facial animation are created automatically. Producing the combination of synthetic speech and facial animation is often referred to as *audio-visual* (or *multimodal) speech synthesis.*

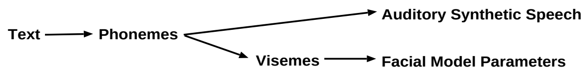


Figure 2.10: Overview of Text-Driven Approach

An overview of the steps taken in text-driven animation is shown in Figure 2.10. Text provided by the user is translated into its corresponding phonetic symbols, which represent the individual speech sounds, or *phonemes*. From the phonemes, the auditory speech can be generated and visemes can be determined. *Visemes* are the visually distinguishable phoneme classes; they correspond to distinct mouth positions formed by the tongue, lips, and jaw. Nitchie [23] has defined 18 visemes associated with the 45 English phonemes; however, there is no standard in defining visemes, and some applications, such as cartoon-like animations, may have smaller sets [8]. There are fewer visemes than phonemes because several phonemes may have the same viseme. For instance, phonemes b, p, and m have the same viseme, which is the closed lips position. Based on the visemes, the facial model's tongue, lips, and jaw can be adjusted to get the desired mouth shape corresponding to the auditory speech.

Several research groups are using this text-driven approach. A few groups have extended Parke's model; these groups include: Kulju et al. [16], Beskow et al. [7] [5] [6], and Cohen and Massaro [9]. Waters and Levergood [29] also have a text-driven approach, but their approach uses muscle-based parameters to create underlying facial expressions, such as happiness or sadness, and does not use parameters to define the lip shapes.

Each of the three groups who are extending Parke's model have different goals and are adjusting Parke's model in different ways. Kulju et al. [16] at Helsinki University of Technology are using an extension of Parke's model to produce a Finnish audio-visual speech synthesizer for studying speech perception, teaching lipreading, and aiding speech therapy. Their extensions to Parke's model include ears and the back of the head, but no tongue. Beskow et al. [7] [5] [6] at KTH in Sweden are extending Parke's model to study people with impaired and normal hearing to see whether supplementing a noisy auditory signal with a synthesized face increases intelligibility. Their implementation of Parke's model includes a tongue. Cohen and Massaro [9] [24] [23] at the University of Santa Cruz have also created an extension of Parke's model for multimodal speech synthesis. The goal of their research is to understand how visual information is used in speech reading, to study how this information is combined with auditory information (as in the McGurk effect where, for example, the facial model mouths "doll", and the audio is "ball", but observers think that model said "wall"), and to improve man/machine communication. Their implementation [9] is a descendant of Parke's model, and it incorporates the code developed by Pearce [23]. Cohen has added many features including: a tongue; texture mapping, which involves "shrink wrapping" an image of real skin onto the model; additional parameters; skin transparency so that the articulations inside of the mouth can be viewed; and control panels to adjust the parameters.

Waters and Levergood [29] created DECface, an audio-visual speech synthesizer. Their goal is to use the animated synthetic face as an interface to allow humans and computers to interact in a more natural manner. Suggested applications are walk-by kiosks, ATM tellers, office environments, and videophones. DECface is a 2D representation of the frontal view with 200 polygons and texture mapping. The implementation includes sliders for six linear muscles so that simple facial expressions can be created. The movement of the lips does not rely on parameters; instead, the movement is topological in that each viseme has a particular topology and "in-between" visemes are generated through interpolation.

2.4.2 Speech-Driven Systems

Speech-driven systems are another approach to speech-synchronized animation. In this approach, a pre-recorded speech sample is analyzed to identify segments of speech. These segments correspond to phonemes, which can be translated into visemes. The goal of the speech-driven system is to take a pre-recorded speech track and create a timed phoneme script that shows the speech phonemes, pauses, and information about when a phoneme begins and ends. From this script, parameters can be derived to control the facial model.

The front end of a speech recognition system can be used for a speech-driven system. Speech recognition systems consist of two parts: acoustic preprocessing and parsing to identify words. In *acoustic preprocessing*, the speech waveform is analyzed to identify speech components such as phonemes and pauses. *Parsing* attempts to identify words corresponding to these speech components. Of these two parts, only the acoustic preprocessing is used in speech-driven systems. However, speech recognition systems may not be the most appropriate choice for producing phoneme scripts since timing information as to when the phonemes occur in the recording is generally not produced, and the accuracy of the acoustic analysis required by the speech recognition systems is not needed for speech-driven systems.

For less accurate acoustic analysis, Lewis and Parke [18] implemented a method based on the linear prediction speech synthesis method. Linear prediction coding (LPC), which can be used to synthesize speech or to re-synthesize natural speech, is based on an excitation signal input to a filter. The excitation signal approximates the acoustic signal produced by the vocal cords. The filter models the vocal tract including the mouth, tongue, and lip positions. Lewis and Parke's implementation is primarily concerned with the filter component of LPC. The filter, represented by filter coefficients, models the positions of the tongue, lip and jaw that create visually distinctive mouth positions (or visemes). Using the filter coefficients, a given speech signal can be classified into a phoneme class which contains a set of phonemes with the same or similar mouth positions. Because the classification involves phoneme sets instead of phonemes themselves, Lewis and Parke's implementation has less accurate acoustic analysis than recognition systems.

In Lewis and Parke's approach, an exact phoneme is not identified for an interval of speech; instead, a visually distinct phoneme class (or viseme) is determined. These visually distinct phoneme classes are represented by 12 *reference phonemes*. These 12 reference phonemes consist of the nine vowels in the words *hate*, *hat*, *hot*, *heed*, *head*, *hit*, *hoe*, *hug*, and *hoot*, and the three consonants *m*, *s*, and *f*. An interval analysis of the prerecorded speech track is compared to a similar analysis of the 12 reference phonemes to determine which phoneme set corresponds to that particular point in time. The results of this analysis and comparison are timed phoneme information, which is saved and used as input to the facial animation system.

2.5 Summary of Animation Techniques

Four animation techniques were discussed in this chapter: image-based key frame, parametric key frame, performance-based, and speech-synchronized. Image-based animation relies on the digitization and interpolation of entire key frame images. The parametric models, which include direct parameterized, pseudo-muscle based and muscle based models, rely on specifying parameter values for each key frame and applying interpolation to the parameter values. This chapter discussed performance-based animation in relation to facial tracking using automatic spot tracking, snakes, and EMG recordings (corresponding to muscle stimulation). Finally, speech-synchronized animation was discussed with respect to text-driven and speech-driven approaches. Table 2.5 summarizes the animation techniques and lists a few advantages and disadvantages of each. The first column identifies the technique, the second gives a brief description of the technique, the third gives the advantages, and the fourth gives the disadvantages.

Technique	Description	Advantages	Disadvantages
Image-based Key Frame	– images of desired key frames are digitized and interpolation is done between the entire images	– realistic results with several key frames	– laborious task of digitizing and interpolating between entire images
Parametric Key Frame	– interpolation occurs on parameter values at desired key frames	– faster than image-based since interpolation occurs on the parameters instead of the entire image – easy to be creative using parameters	– must have a good choice of parameters – changes to several parameters simultaneously may yield unpredictable results
Performance-based	– human actions drive the animation	– simple to obtain animations quickly – can catch subtle face actions	– difficult to track actual facial movement without invasive markers or special lighting conditions
Speech-Synchronized	– the animation is connected to auditory speech – usually consists of breaking down the text or audio into phonemes or speech sounds	– animation and sound are directly connected so that changes in the speech immediately affect the animation	– may inherit the problems associated with text-speech generation and speech recognition – in text-driven, the audio is a synthesized voice

Table 2.5: Advantages and Disadvantages of the Major Animation Techniques

2.6 Animation Technique Used in this Research

The animation technique used in this research can be classified as performance-based, since human actions drive the animation. The human actions are recorded as X-ray microbeam data that give the positions of the lips, jaw, teeth, and tongue. An advantage of this approach over other performance-based approaches is that the movement of the tongue inside the mouth is captured and used to drive the animation. In other techniques, such as those performance-based techniques that use video camera recordings, the tongue movement cannot be monitored.

A modified form of Parke's model was used. Parke's model was chosen because of its simplicity and low computational complexity (when compared to the muscle based models), because the source code was easily accessible, and because several researchers interested in the correspondence between the auditory signal and the animation are using Parke's model [9] [6] [16]. To modify Parke's model for use in a performance-based approach with X-ray microbeam data, the functionality of certain parameters were disabled, new parameters were created, and new jaw rotation and lip movement procedures were developed. With these adjustments, the two-dimensional X-ray microbeam data could be used to drive the three-dimensional Parke's facial model.

Kinematic animation refers to "motion specification in terms of positions, velocities, and acceleration over time, neglecting the forces and torques that actually cause the motion" [1]. The work described in this document falls into the category of kinemetic animation. Motion is directly generated by a series of frames set according to the X-ray microbeam data. No interpolation is required since each frame, in essence, becomes a "key" frame.

Chapter 3

Inherited Coordinate Systems and Data

The CASSI (Computer Animated Speech SImulator) software system combines two things: Parke's model and X-ray microbeam data. Parke's model is a three dimensional facial model with parameters to alter the size and shape of the various parts of the face. The X-ray microbeam data is the two dimensional recording of the side view movement of pellets placed on the speaker's tongue, lips and jaw while performing speech related tasks. These two things form the foundations of the CASSI system and are discussed, in further detail, in this chapter. Section 3.1 describes the topology, parameters, and jaw rotation procedure inherited from Parke's model. Section 3.2 describes the X-ray microbeam data including details on the subjects and the tasks performed, tracking and mistracking pellets, the coordinate system, post processing, and the resulting data.

3.1 Parke's Topology and Parameters

The CASSI system is based on the animated facial model implemented by Fredric I. Parke. Code for this model was obtained from: "http://www.crl.research.digital.com/publications/books/waters/Appendix2-/ap2.html". For animation and interface ideas, code from Andrew Marriot, who extended Parke's implementation, was used. Andrew Marriot's code was obtained from: "http://mambo.ucsc.edu/psl/Fascia/" and contained the same underlying topology, vertices, and parameters as Parke's model. The major difference is that Marriot's code included a GUI interface, which is briefly discussed in Appendix A. The topology and parameters of Parke's model are described in further detail in this section. In addition, Parke's original jaw rotation procedure is described as it forms the foundations of the jaw rotation procedure in CASSI 1.0.

3.1.1 Topology and Coordinate System

In Parke's model, the surface of the face is approximated by a series of polygons. Each polygon is defined by three or four 3D points (or vertices). The way in which these vertices are connected is the polygon topology. Figure 3.1 shows Parke's polygon topology from the side and front views. The numbers in the figure correspond to specific vertex numbers. For example, vertex 1 refers to the point at the centerline base of the neck. In this coordinate system, X is forward, Y is to the face's left, and Z is up.

Three input files are used to create the 3D surface of the face: *st1.pts*, *st2.pts*, and *stt.top*. The two files: *st1.pts* and *st2.pts* are vertex files containing the 3D (X,Y, and Z) coordinates of points of the face. *st1.pts* is used to define a baseline of face vertices, while *st2.pts* is used to define extreme values for vertices which are computed using interpolation. The third input file, *stt.top*, defines the polygon topology.

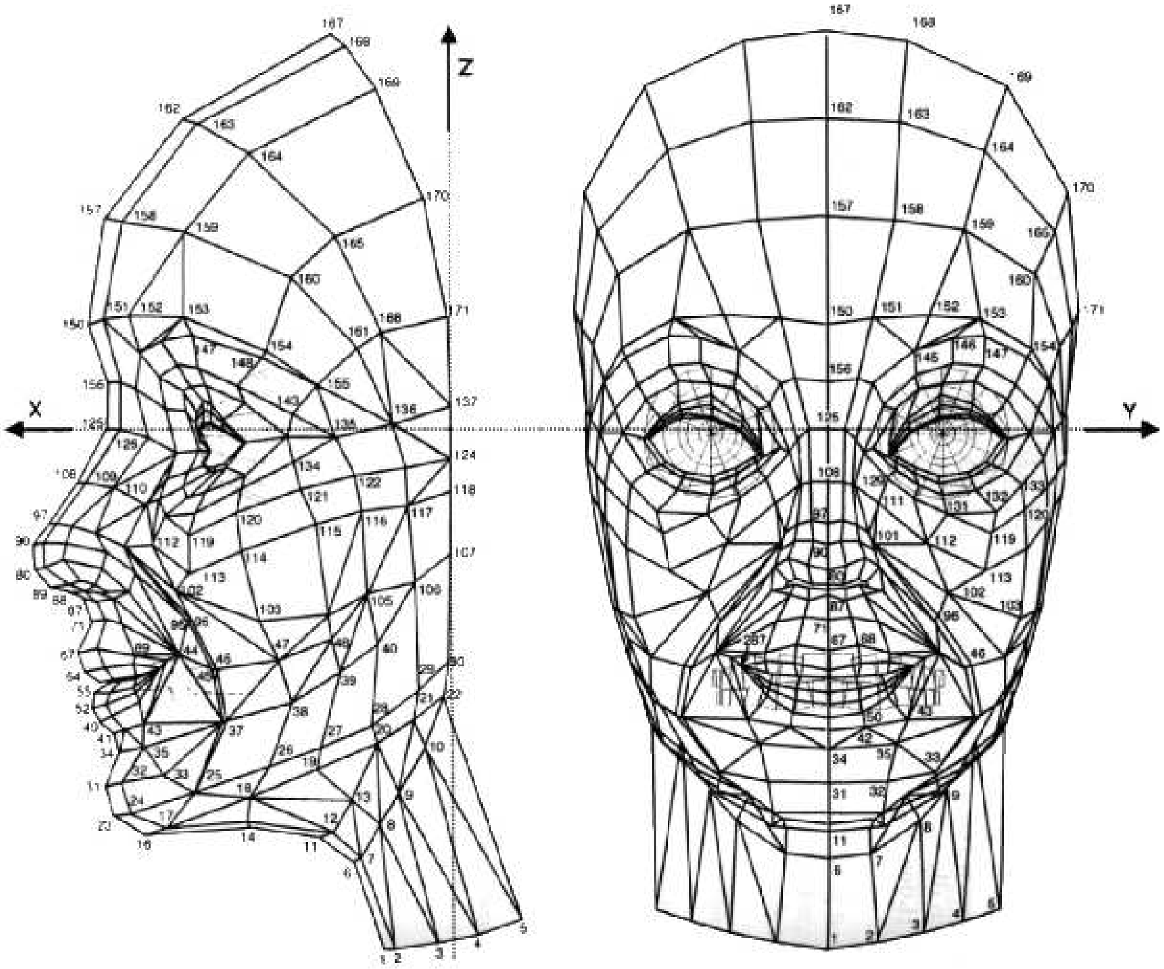


Figure 3.1: The Polygon Topology Used in Parke's Implementation (taken from [23])

Both vertex files (*st1.pts* and *st2.pts*) have the following format:

```
{description of file}
{number of vertices}   {number of vertices in file}
{vertex number} {x coordinate} {y coordinate} {z coordinate}
{vertex number} {x coordinate} {y coordinate} {z coordinate}
{vertex number} {x coordinate} {y coordinate} {z coordinate}
                              .
                              .
                              .
```

A short excerpt from *st1.pts* is provided in Figure 3.2. The first line, "st1.pts - first vertices set FIP 10 May 90", describes the file. The second line indicates that there will be 287 vertices all together, but only 275 vertices are defined in *st1.pts*. In the lines that follow, the vertices are defined. For example, the third and fourth lines define two vertices along the base of the neck: vertex 1 with coordinates (45, 0, -594) and vertex 2 with coordinates (28, 50, -585). These vertices can be seen in Figure 3.1. *st2.pts* has the same format as *st1.pts* but contains a second set of coordinate values for vertices that are interpolated. For instance, the position of the eyelid is determined by the interpolation of two extreme sets of coordinate values (one for open eyelid and one for closed eyelid). If, for example, the interpolation parameter is 50%, then the eyelid will be halfway between open and closed. The coordinates for the open eyelid are defined in *st1.pts* and the coordinates for the closed eyelid are defined in *st2.pts*.

```
st1.pts - first vertices set FIP 10 May 90
287     275
1       45      0       -594
2       28      50      -585
3       -8      105     -565
4       -45     147     -542
5       -90     187     -518
6       100     0       -450
7       93      45      -445
8       70      97      -412
9       52      127     -380
10      24      158     -330
11      138     0       -425
12      123     45      -420
13      103     85      -386
14      210     0       -415
15      215     50      -410
16      326     0       -423
17      302     60      -418
                .
                .
                .
```

Figure 3.2: Excerpt from *st1.pts*

The polygon topology file (*stt.top*) has the following format:

```
{description of file}
{number of polygons defined in stt.top}
{material number} {vertex number} {vertex number} {vertex number} <vertex number>
{material number} {vertex number} {vertex number} {vertex number} <vertex number>
{material number} {vertex number} {vertex number} {vertex number} <vertex number>
                                              .
                                              .
                                              .
```

```
stt.top - simple face topology file FIP 10 May 90
254
4       222.1   212.1   190.1
4       223     224     225     226
4       227     228     229     230
4       231     232     233     234
4       235     236     237     238
4       239     240     241     242
4       243     244     245     246
4       247     248     249     250
4       251     252     253     254
4       255     256     257     258
4       259     260     261     262
4       263     264     265     266
4       267     268     269     270
4       271     272     273     274
4       275     276     277     278
4       279     280     281     282
4       283     284     285     286
1       1       6       7       2
1       7       8       2
1       2       8       3
                .
                .
                .
```

Figure 3.3: Excerpt from *stt.top*

The first two lines describe the file and each of the subsequent lines describe a three or four-sided polygon. Each line describing a polygon has a material number followed by either three or four vertex numbers (the fourth vertex is shown above in "<>" to indicate that it is optional). The material number refers to the color and lighting properties that have been defined for the specific parts of the face. For instance, material 1 is the flesh, material 2 is the lips, material 3 is the eyelash, and material 4 is the teeth. Figure 3.3 shows an excerpt from *stt.top*. The first line, "stt.top - simple face topology file FIP 10 May 90", describes the file. The second line indicates that there are 254 polygons defined in *stt.top*. The third line and subsequent lines describe the polygons. All lines starting with 4 describe polygons associated with the teeth, and all lines starting with 1 describe polygons associated with the flesh. The third line, "4 222.1 212.1 190.1", defines a three-sided polygon for the teeth (material 4). The vertex numbers are 222, 212, and 190 which have associated x, y and z coordinates (defined in *st1.pts*). The ".1" refers to the normal used to resolve multiple normals for vertices which occur along creases. All other teeth polygons in *stt.top* have a material number of 4 and are defined by four vertices. The first polygon for the flesh is defined by the line "1 1 6 7 2". The first 1 corresponds to the material number of flesh, and the other numbers (1, 6, 7, and 2) refer to the vertex numbers which make up a four-sided polygon. This polygon can be seen in Figure 3.1 as the base centerline portion of the neck with the vertices labelled: 1, 2, 6 and 7.

In summary, the vertex files (*st1.pts* and *st2.pts*) define the x, y, and z coordinates for specific vertex numbers, and the topology file specifies the interconnections between these defined vertices. The model assumes that the face is symmetric from left to right. Thus, vertices are only defined for the face's left side, since vertices for the right side are assumed to be its mirror image.

3.1.2 Parameters

P #	Start Value	Description	Effect of Change
1	.9	eye opening (0.0−1.0)	1−eyelids open, 0−eyelids closed
2	.5	eyebrow arch (0.0−1.0)	1−eyebrows upside down V shape, 0−eyebrows slope down(center to edge)
3	.0	eyebrow separation (0.0−25.0)	25−eyebrows separated, 0−eyebrows close together
4	.0	jaw rotation (0.0−20.0)	20−jaw open, 0−jaw closed
5	1.0	eyelid Y scale	1.1−eyelid stretches horizontally, 0.5−eyelid half width (horizontally)(only pupils can be seen)
6	1.0	eyelid Z scale	1.5−eyelid widens (extreme open eyes), 0.5−eyelid half width (like squint)
7	1.0	head X scale	2−stretches the face (front to back), 0.5−flattens the face (front to back)
8	1.0	head Y scale	2−widens the face, 0.5−thins the face
9	1.0	head Z scale	2−lengthens the face (top to bottom), 0.5−squishes the face (top to bottom)
10	1.0	nose tip Y scale	2−makes tip of nose twice as wide, 0.5−makes tip of the nose thin
11	1.0	nose bridge Y scale	2−widens the bridge of the nose, 0.5−thins the bridge of the nose
12	1.0	mouth Y scale	2−lips stretch horizontally, 0.5−lips shrink horizontally (edges poke through teeth)
13	1.0	mouth interpolation (0.0−1.0)	1−lips are slightly turned up(smile), 0−lips are slightly turned down (frown)
14	.0	mouth X offset	20−lips are moved forward, -20−lips are moved back (run into teeth)
15	.0	287 Y offset	no apparent effect
16	.0	mouth corner X offset	10−corners move forward, -10−corners move back
17	.0	mouth corner Y offset	10−corners move away from each other, -10−corners move towards each other
18	.0	mouth corner Z offset	10−mouth corners move up, -10−corners move down
19	1.0	jaw Y scale	2−jaw is twice as wide, 0.5−jaw is half as wide
20	1.0	cheek Y scale	2−cheeks stick out twice as much, 0.5−cheeks indent into face
21	.0	lower lip 'f' tuck	10−lower lip moves down and away from teeth, -10−lower lip moves up and towards teeth
22	15.0	raise upper lip	30−upper lip moves up (large gap between lips), 0−upper lip moves down (lips are pursed)
26	.0	nose tip X offset	10−nose tip moves forward, -10−nose tip moves back
27	.0	nose tip Z offset	10−nose tip moves up, -10−nose tip moves down
28	.0	eyeball X offset	10−eyeballs move forward out of socket, -10−eyeballs move back into socket
29	.0	eyeball Y offset	10−eyeballs move away from center of face, -10−eyeballs move toward center face (look at nose)
30	-15.0	eyeball Z offset	0−eyeball move up (like looking up), -30−eyeballs move down (like looking down)
31	.0	chin X offset	10−chin moves forward, -10−chin moves back
32	.0	chin Z offset	10−chin moves up, -10−chin moves down
35	1.0	chin to mouth scaling	2−chin becomes twice as long, 0.5−chin becomes half as long
36	1.0	chin to eye scaling	2−stretches the face (between eyes and chin), 0.5−squishes the head (between eyes and chin)
37	1.0	eye to forehead scaling	2−forehead stretches up (cone head), 0.5−forehead squishes down from eyebrows up
38	.0	eyelid X offset	20 eyelid and eyeball move forward, -20 eyelid and eyeball move back into face
39	.0	eyelid Y offset	10−moves eyes farther apart, -10−moves the eyes closer together
40	.0	eyelid Z offset	10−moves eyes up, -10−moves the eyes down
41	.4	pupil fraction	0−no pupil (all green), 1−mostly pupil with thin rim of green
42	.85	fringe fraction	1−no rim around the iris, 0.1−large rim, really small pupil
43	.4	iris fraction	1−iris covers eye (no whites), 0−no iris (all white)
44	70.	eyeball radius	80−eyeball is too big (pokes out of socket), 35−eyeball is too small for the socket
45	80.	eyelid radius	100−eyelid moves forward away from the eyeball, 75−eyelid moves back into the eyeball
46	0.0	growth factor	0.1−increases face (mostly in cheeks and jaw)−looks older, -0.1−decreases face (in cheeks/jaw)
47	0.0	teeth X offset	20−teeth move forward (poke through lips), -10−teeth move back
48	38.0	teeth Z offset	70−teeth move up, 0−teeth move down
50	1.0	smoothness (1.0−0.0)	1−smooth surface (few edges viewable), 0−faceted (can see polygon edges)

Table 3.1: Parameters Provided in Parke's Implementation

In Parke's original implementation, adjustments can be made to the face by manually changing the values of several defined parameters. There are two broad categories of parameters (as described by [22]): conformation and expression. The first category of parameters control the conformation or structure of the individual face. These parameters relate to aspects of the face that vary from individual to individual. Examples of conformation parameters are: jaw width; eyelid, eyeball, and iris size; the position and separation of the eyes; chin, forehead, cheek, and cheek bone shape; nose length and the width of the bridge and end of the nose;

and chin and forehead scale. The second category are the expression parameters which relate to displaying emotion or to generating lip and jaw positions used for speech. Examples of expression parameters include: pupil dilation, eyelid opening, eyebrow position and shape, the direction in which the eyes are looking, jaw rotation, width of the mouth, position of the upper lip, and positions of the corners of the mouth.

Table 3.1 provides a list and brief description of all the conformation and expression parameters inherited from Parke's implementation. The first column provides the parameter number. The second column contains the starting value of the parameters for the initial display of the face. The third column is a brief description of the parameters; where "(*value* − *value*)" in some descriptions indicate that the parameter values should fall in that range. The fourth column describes the effect of changing the parameters from the starting value. In most cases, we chose parameter values which were higher and lower than the original starting value. For each value, we then examined the face for changes. In most cases, the modified face is peculiarly shaped, and the purpose of the parameter became clear. For example parameter 4 (jaw rotation) is 0 in its starting location, and the jaw is closed. If the value is changed to 20, the jaw is rotated to a wide open position. As another example, parameter 37 (eye to forehead scaling) has a starting value of 1.0, with a normally proportioned forehead. By adjusting the value to 2.0, the face changes to have a long forehead, reminiscent to that of a cone head character. Similarly, by changing the value to 0.5, the forehead is squished from top to bottom, also resulting in an abnormally shaped face.

3.1.3 Jaw Rotation

Jaw rotation, as implemented by Parke, depends on the rotation of vertices around a fixed point corresponding approximately to the joint of the jaw. Because the CASSI system uses a modified version of Parke's jaw rotation procedure, it seems appropriate to discuss Parke's original implementation. The following two sections discuss Parke's jaw rotation in relation to the underlying formulas for rotation and the details of the rotated vertices.

Rotation Formulas

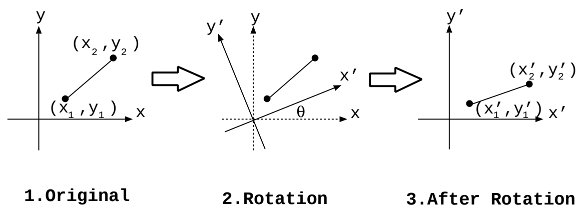


Figure 3.4: Rotation of the Axes around the Origin

Parke's original jaw rotation procedure is based on the formula for the two dimensional rotation of the axes around the origin. Figure 3.4 shows the rotation of the axes around the origin and the effect such a rotation has on a line segment denoted by (x_1, y_1) and (x_2, y_2). The angle of rotation is θ and the resulting line segment after rotation is denoted by (x'_1, y'_1) and (x'_2, y'_2). The formulas for such a rotation are the following [2]:

$$x' = x\cos\theta + y\sin\theta \tag{3.1}$$

$$y' = y\cos\theta - x\sin\theta \tag{3.2}$$

Sometimes, we may want to know the result of rotating around some arbitrary point P, denoted by (p_x, p_y). To determine this result, we use the following three transformations:

1. Translate such that P is at the origin
2. Rotate (using Equations 3.1 and 3.2)
3. Translate such that P returns to its original position

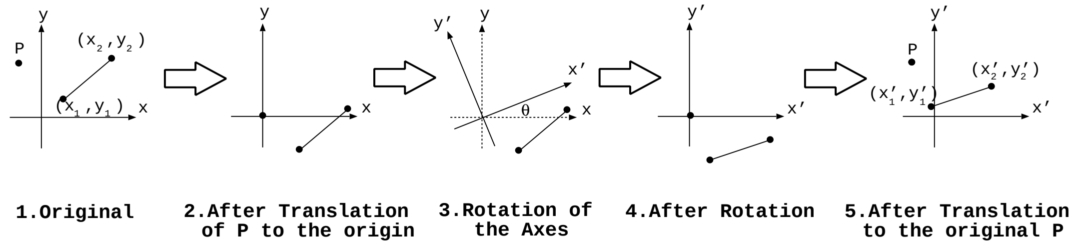


Figure 3.5: Rotation Around a Fixed Point

These three operations are shown in Figure 3.5. The first coordinate system shows the original line segment, denoted by (x_1, y_1) and (x_2, y_2), and the rotation point, P. The subsequent coordinate systems show the result of performing the three transformations. The final coordinate system shows the resulting line segment, denoted by (x_1', y_1') and (x_2', y_2'), with the point of rotation, P, in its original location. Formulas derived from these three transformations are as follows:

$$x' = (x - p_x)\cos\theta + (y - p_y)\sin\theta + p_x \tag{3.3}$$

$$y' = (y - p_y)\cos\theta - (x - p_x)\sin\theta + p_y \tag{3.4}$$

where the first translation of point P to the origin is shown by $(x - p_x)$ and $(y - p_y)$, the rotation is from Equations 3.1 and 3.2, and the second translation of the origin to P is shown by the addition of p_x and p_y to the end of the formulas.

Rotation of Parke's Jaw

Equation 3.3 and 3.4 are used in Parke's procedure for jaw rotation. The procedure rotates the jaw and surrounding vertices around a fixed point (vertex 107) which roughly corresponds to the joint of the jaw. Equations 3.3 and 3.4 are rewritten by substituting the z for y (to correspond to Parke's coordinate system), and by substituting $vertex[107]_x$ and $vertex[107]_z$ for p_x and p_z:

$$x' = (x - vertex[107]_x)\cos\theta + (z - vertex[107]_z)\sin\theta + vertex[107]_x \tag{3.5}$$

$$z' = (z - vertex[107]_z)\cos\theta - (x - vertex[107]_x)\sin\theta + vertex[107]_z \tag{3.6}$$

Parameter 4 determines the value of θ used in Equations 3.5 and 3.6. Its value, which is between 0 and 20, corresponds to the angle of rotation (θ). For instance, a value of 0 rotates the jaw 0^o, forming the closed jaw, and a value of 20 rotates the jaw 20^o, forming the open jaw. Choosing values between 0 and 20 for parameter 4, forms jaws intermediate between open and closed.

A rotation zone consists of a set of points or sets of points that have an equal angle of rotation. This angle of rotation is specified as a fraction of the angle of rotation for the jaw itself (the component with the most motion). For instance, vertices forming the jaw are rotated by θ, but vertices forming the part of cheek which stretches with the jaw movement are rotated by a fraction of θ. The groups of vertices and their corresponding fraction of θ are shown in Figures 3.6 and 3.7 and are summarized by Table 3.2.

Figure 3.6 shows the side view of the jaw and Figure 3.7 shows the side view of the lips. Both figures have vertex numbers and shaded regions which correspond to the different rotation zones. The zones are defined by the different portions of θ used for rotation and are labelled "1st" through "5th". For instance, in Figure 3.6 vertices labelled 44, 45, 46, 47 and 48, corresponding to the cheek, are shown in a shaded region denoted by "5th zone"; these vertices are rotated by $.35\theta$ and, thus, have little movement compared to the jaw vertices.

Figure 3.6 shows the 1st, 3rd, 4th and 5th rotation zones for the cheek, jaw and neck. The 2nd zone is not included because it consists of the set of vertices for the lower teeth (vertices 255 to 286) and a set of lip vertices (50, 53, 56, and 59); both sets could not be included due to lack of space. Although the details of the lips and nose were not shown in Figure 3.6, the outlines of these were shown by the sampling of vertices 49 through 70 for the lips and 80 through 101 for the nose.

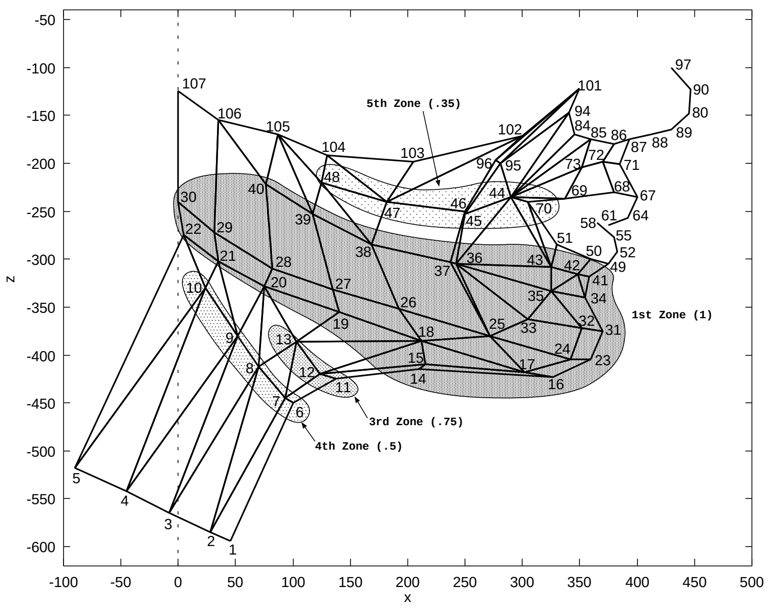


Figure 3.6: Parke's Rotation Zones of the Cheek, Jaw and Neck (Side View)

Figure 3.7 shows the side view of the lips. The upper lip is shown as vertices 61 to 69, the lower lip as 49 to 60, and the corners as 70 and 287. Figure 3.7 shows the rotation zones for the lips exclusively. Because the 4th zone does not have any set of lip vertices, it is not included in this figure.

Rotation Zone	Fraction of θ	Vertices	Description	Values Adjusted
1st	1	14 to 43	jaw	x and z
		49, 52, 55, 58	centerline of lip	x and z
2nd	.9	50, 53, 56, 59	second set of lip vertices	x and z
		255 to 286	lower set of teeth	x and z
3rd	.75	11, 12, 13	underneath chin	z
		51, 54, 57, 60	third set of lip point	x and z
4th	.5	6 to 10	neck vertices	z
5th	.35	44 to 48	lower cheek	x and z
		70	corner of lip	x and z
		287	inside corner of lip	x and z

Table 3.2: Summary of Vertices Rotated to Create Jaw Rotation in Parke's Implementation

Table 3.2 summarizes the set or sets of vertices in a rotation zone and the fraction of θ used to rotate them. The first and second columns hold respectively the rotation zone label and the corresponding fraction of θ, the third column lists sets of vertices, the fourth column describes these sets, and the fifth column indicates whether both x and z are rotated or z alone. For example, neck vertices 6 to 10 are in the 4th rotation zone, and only the z coordinate values are rotated by 0.5. Although a corresponding diagram is not shown, Table 3.2 indicates that the lower set of teeth (vertices 255 to 286) is rotated by $.9\theta$.

3.2 X-Ray Microbeam Data

The X-ray microbeam (XRMB) data were provided by the University of Wisconsin. The major goal of the XRMB research at the University of Wisconsin was to gain a better understanding of the relationship between articulatory movement (movement of body parts such as the tongue, lips, and jaw) and the resulting

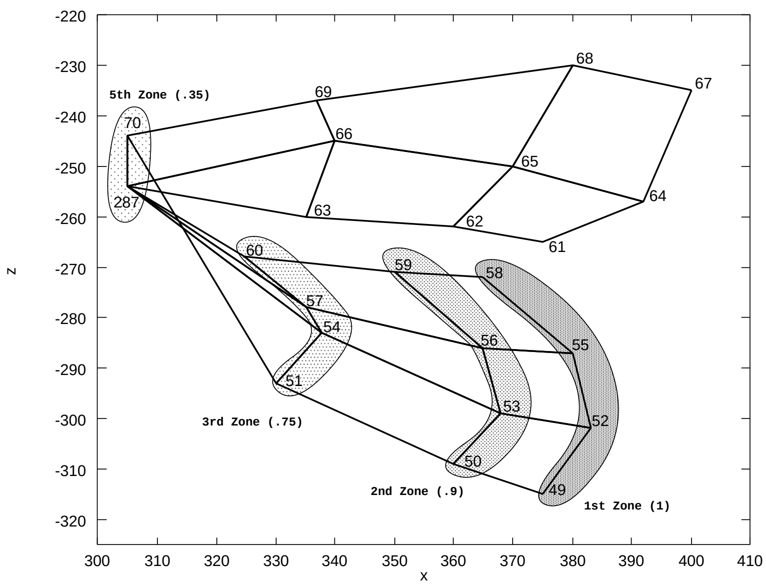


Figure 3.7: Parke's Rotation Zones of the Lips (Side View)

speech wave. For this, a method of capturing the movement of the tongue, in particular, was needed. Regular video could not be used because the motion of the tongue is mostly hidden behind the lips and teeth. The chosen method involved X-ray microbeams tracking the motion of 2-3 mm diameter gold pellets attached to the tongue, jaw and lips of a person directed to perform speech related tasks. Instead of obtaining a full X-Ray of the head, microbeams were used to sample only the areas where the pellets were expected. This limited the speakers' exposure to ionizing radiation [30].

In the sections that follow, the details of the XRMB database are discussed including: the subjects and tasks, the process of tracking the pellets, mistracking, the coordinate system, the placement of the pellets, the post-processing, and the resulting data.

3.2.1 Subjects and Tasks

The XRMB database contains 57 subjects (32 females and 25 males) performing up to 118 tasks. The subjects' average age was 21.1. These subjects were recruited by advertisements from the campus of the University of Wisconsin-Madison and from the surrounding city. Some criteria had to be met by the subjects: the subjects had to be largely free of dental fillings, which may be mistaken for pellets; they had to pass a screening test for hearing, speaking and reading ability; they had to be in good general health; and they had to be willing to participate (candidates were told about the risks and discouraged from participation if they had any misgivings about the experiment).

Data collection included head measurements, jaw measurements, and teeth measurements along with the collection of the pellet coordinates and associated speech data. Measurements of the subjects head and jaw shape were recorded for reference. The cavity created by the teeth was determined by creating plaster models of the upper and lower dental arches. To collect pellet coordinates, eleven pellets were attached in the following places: three to the head as reference or fiducial markers, one to the upper lip and one to the lower lip, four on the tongue surface and two on the mandible (lower jaw). Recordings of the sound waves and of the associated pellet coordinates were made as speakers performed speech related tasks. For some people, the pellets in their mouth had a perceptible effect on their speech while for others, they had little effect, especially towards the end of all of the tasks.

Each subject performed up to 118 speech related tasks. These tasks included: number names, phrases made from number name sequences (for example: "9739286"), oral motor tasks (for example: jaw wagging, maximal tongue and lip protrusion, and swallowing), citation words, citation sVd's (for example: "side", "sad", and "said"), isolated vowels, vowel sequences, VCV's (vowel consonant vowels), and sentences (including some DARPA/TIMIT material). During the recordings of these tasks, the subject's head was supported from the rear by a head rest, but was otherwise unrestrained. Because of the nature of the XRMB system, speakers were encouraged to remain as still as possible and were able to monitor their position by viewing a mirror reflection of a low intensity laser beam projected onto a specifically marked position on the forehead.

3.2.2 Tracking and Mistracking Pellets

The pellets are tracked by a narrow beam of high energy X-rays directed by a computer and scanning roughly a 6 mm square area where one particular pellet is expected to be. This expected location is determined by using current and previous positions of a pellet to predict its future position along its trajectory. The position of the pellet is assigned when it falls in such a scan area and produces a recognizable "shadow". For any discrete moment of time, the system assigns coordinates to the pellets based on the centroids of their respective shadows. The cycle of local scan, recognition, and prediction is repeated for each pellet at a combined rate of approximately 700 times/second.

Unfortunately, for one discrete moment, only one pellet's coordinate values can be recorded. The XRMB is designed to share the tracking rate of approximately 700 times/second amongst all 11 pellets. This rate does not need to be divided equally since certain pellets move more quickly than others. Table 3.3 shows the schedule on which the sample rate of each pellet was based. The first column gives the pellet type, the second column gives the pellet name where the "(a)" stands for a letter; for instance, the two mandibular (or lower jaw) pellets are MANi and MANm. The third column is the number of pellets of that type. The fourth column is the sampling rate for each pellet of that type, and the final column is the running total of the sampling rates needed for the pellet type in each row. This schedule of sample rates was used as a guideline and was adjusted if speakers had uncommonly quick or jerky movements or if the speaker had less than 11 pellets. To coordinate all pellet samples to a synchronized rate of 160 samples/second, post processing was used to interpolate and re-sample the pellet positions.

Pellet type	Pellet name	N	Nominal sampling rate	Running Total
reference	MAX(a)	3	40 samples/second each	120 samples/second
mandibular	MAN(a)	2	40	200
upper lip	UL	1	40	240
lower lip	LL	1	80	320
ventral tongue	T1	1	160	480
mid-tongue	T2, T3	2	80	640
dorsal tongue	T4	1	80	720

Table 3.3: Schedule of Pellet Rates (Modified from [30])

During tracking, current and previous positions of a pellet are used to predict its future position. At the start of the recording session for the subject, however, no previous position exists. To determine the location of the pellets before tracking, an *initialization scan* is used. This involves a rapid sweep of the X-ray beam over the central 15x15 cm portion of the image field. The result looks something like a fuzzy radiographic image that can then be displayed to a system operator on a graphics monitor. The operator visually estimates the pellet locations and enters them into a file to direct the computer on where to aim the beam for any given pellet. This initialization is usually good for several records (corresponding to the tasks). However, if the speaker position changes too much between records, the initialization scan is repeated.

During successful tracking, a pellet is followed by a local X-ray scan determined by the pellet's earlier found and current positions. The pellet's location is identified by a recognizable shadow on the scan, and its coordinates are determined by the centroid of this shadow. During this tracking process, sometimes pellets are momentarily lost. At those times, pellets are said to have been *mistracked.*

During mistracking, no shadow, or an unrecognizable shadow may occur. The XRMB system responds to this by displaying a "not found" error message, and by incrementing a variable which holds the total number of "not found" samples per pellet, per record. It also responds in two additional ways: by returning to the location where the pellet was first found in the record in hopes that it will pass nearby and be captured again, and by returning a false coordinate value for the pellet.

There are a few possible causes for mistracking. First, the change in pellet velocity between previous and current locations may be so great that the estimated location is wrong. Second, there may be insufficient contrast between the signal shadow cast by the gold pellet and the background shadows of tissue, bone, teeth, and fillings so that the pellet is unidentifiable. Third, the scan may start to follow another pellet so that two scans meant for two different pellets start to follow the same pellet. Finally, the scan may choose the shadow cast by something pellet-like (for example, a dental filling or unusual dense tissue) and follow it.

3.2.3 XRMB Coordinate System and Pellet Positions

Figure 3.8 shows the coordinate system for the XRMB database. The coordinate system is placed on the midsagittal plane, which is the plane resulting from a vertical, symmetric slice made directly down the center of the face. Because no two people's heads have the same size and shape, a special way of defining the coordinate system is used. First, the origin is defined to be between the tips of the central maxillary incisors (top front teeth). The x-axis corresponded to the intersection of the midsagittal plane and a second plane, referred to as the maxillary occlusal plane (MaxOP). This plane, shown in Figure 3.8, runs horizontally along the tips of the top set of teeth, in particular, along the tips of the two top front teeth and at least two other top teeth on opposite sides of the mouth. The midsagittal plane itself is assumed to be normal to the MaxOP, containing the line passing midway between the gap made by the top front teeth. The y axis is normal to the MaxOP and intersects that plane at the origin. Positive x movement is out of the mouth, and positive y movement is toward the roof of the mouth.

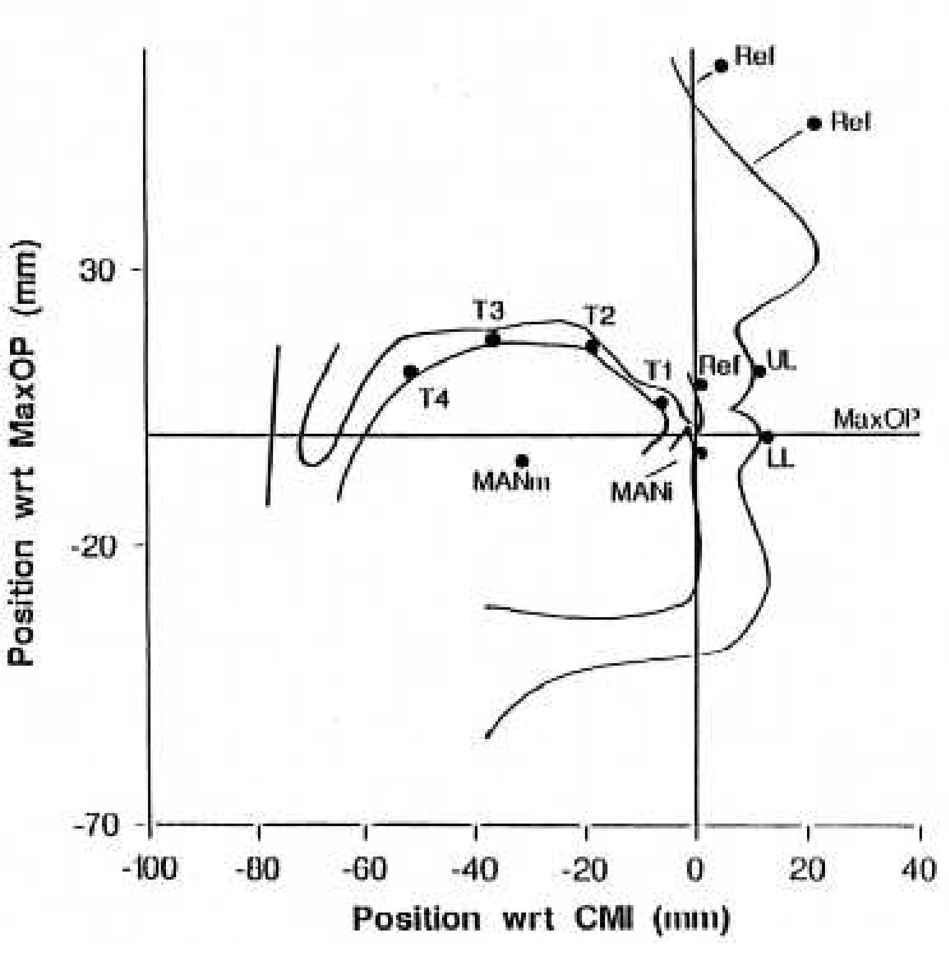


Figure 3.8: XRMB Coordinate System and Approximate Pellet Placement (taken from [30])

Figure 3.8 shows the y axis labelled as "Position wrt [with respect to] MaxOP(mm)" and the x axis labelled as “Position wrt CMI [central maxillary incisors] (mm)”. Each speaker has his or her own coordinate system depending on the location of their central maxillary incisors (top front teeth) and the resulting maxillary occlusal plane (defined by the top set of teeth). Having this coordinate system with an anatomically-based reference frame common to all speakers makes the data easier to describe and interpret across speakers, and provides a more intuitive sense of direction since “up” is toward the top of the head, and “forward” is toward the face front.

There are 11 pellets shown in Figure 3.8: three reference pellets (Ref), four tongue pellets (T1, T2, T3, and T4), two jaw pellets (MANm and MANi), and two lip pellets (UL and LL). The three reference pellets were used to remove global head motion from the motions of the remaining eight pellets. The highest reference pellet was placed on a short stand-off post glued to the bridge of the nose. A second reference pellet was also glued in the vicinity of the nose either lower along the bridge of the nose or attached to an arm projecting from a snug-fitting pair of eyeglass frames. The third reference pellet was attached to the outside of the top front teeth in the pocket between the two central teeth where the gums and the teeth meet. Four pellets were attached along the central groove of each speaker's tongue. The most forward tongue pellet (T1) was placed roughly 10 mm behind the tip of the extended tongue. The farthest back (T4) was place about 60 mm behind the tip, as far back as the speaker could tolerate without gagging. The two middle tongue pellets (T2 and T3) were placed so that the distance between the front and rear-most pellets was divided into three roughly equal segments. Two jaw pellets represent the motion of the jaw. One pellet, MANi (mandibular incisor), was glued to the outer surface of the central incisors (bottom front teeth) in the pocket formed by the gap between the teeth where the gums and the teeth meet. The second jaw pellet was attached farther back on the jaw in the vicinity of the area between the first and second mandibular (lower jaw) molars, either where the teeth and the gums meet, or on the gum itself. The final pellets were the lip pellets; one was attached to the upper lip (UL) and the second was attached to the lower lip (LL). Both were glued to the outside of the lips at the upper and lower edges of the lips.

3.2.4 Post-processing

Rectangular image-plane coordinates are assigned to the evaluated pellet center during real-time tracking. In order to maximize accuracy and to standardize these coordinates the following post-processing steps are executed: (1) make target to image-plane corrections, (2) re-sample the pellets to equal time intervals, (3) translate the coordinates from “machine space” to “head space”, and (4) evaluate the position of the head. The first step (making target to image-plane corrections) involves removing the distortion of the XRMB system cylindrically curved tungsten target, where the X-ray beam originates; and numerically scaling pellet position data from the image-plane to life size. The second step involves interpolation, using piece-wise continuous smoothing splines; and re-sampling of the pellets to a uniform rate of 160 samples/second for all pellets at common times. The next step involves translating the pellet coordinate values from “machine space” to “head space”. This step involves removing the rotational and translational components of the head motion by using the reference pellets; and translating and rotating to establish the coordinate system origin as the tip of the top, front teeth (central maxillary incisors) and the x and y axes as described in Section 3.2.3. Head motions other than pitching rotation (about the axes normal to the midsagittal plane) and translation relative to the axes lying on the midsagittal plane are more difficult to remove. Westbury [30] discusses the calculated magnitude of the errors due to “off plane” and “off-normal orientations”. “Off plane” errors occur when the head is shifted closer to or farther from the XRMB target. “Off-normal orientations” occur when the head is moved from side to side or from shoulder to shoulder. The final step in the post-processing involves treating scale changes associated with these head motions as simple translations along the XRMB system z-axis. To do this, scalar multiplication by the actual distance between two reference pellets over the current distance between the same two reference pellets is applied to each pellet-position sample in each record.

3.2.5 Resulting Data

Of the entire database of 57 subjects, we were provided with a subset of 16 subjects denoted by: JW11, JW12, JW15, JW16, JW18, JW19, JW21, JW24, JW25, JW27, JW29, JW32, JW40, JW41, JW45, and JW502. This data was provided to us on a compact disk with a sub-directory for each speaker. Ideally, in each subject's directory, there should be 118 records (one for each task); however, for some speakers, some tasks were recorded more than once in response to mistracking or other acquisition flaws, and for other speakers, records were damaged and lost during acquisition, due to system errors or deleted because they were judged to be unacceptably flawed or uninterpretable. Each record has an audio file (**.acc*), a throat accelerometer file (**.tcc*), and an X-ray pellet coordinates file (**.xyd*). Each subject has at least three additional files: *Mis.dat* which provides guidelines on which pellets were mistracked and where; and two files defining the vocal tract boundary outlines, *Pal.dat* which defines the roof of the mouth or palate, and *Pha.dat* which defines the back of the mouth or pharynx. In a separate directory from the database, an executable, *sp.bat*, existed which made it possible to view the speech wave and the corresponding 2-dimensional XRMB pellet coordinate values for one task. With this executable, pellets were shown as moving ticks with the the tongue pellets connected by a curved line and the jaw pellets connected by a straight line. In addition, for reference, the palate and pharynx were drawn as curved and straight lines respectively.

As mentioned above, there are three files for each record: the audio recording file, the throat accelerometer file, and the X-ray pellet coordinates file. Both the audio recording and the throat accelerometer files are waveform samples that have been compressed. The audio recording is obtained from the output of a microphone positioned at mouth level. The throat accelerometer represents the vibration of the neck wall and has been recorded in anticipation of future interests in LPC analysis of the speech acoustic wave. The X-ray pellet coordinates file **.xyd* is the primary focus of this implementation. These files contain pellet data in ASCII format. The data is sampled every 6.866ms, and contains the x and y coordinates for each pellet relative to its previous location. For our purposes, we used an executable, *unxyd.exe*, provided with the XRMB data to convert the **.xyd* files to **.txy* files. The **.txy* files are different from the **.xyd* files in two major ways: (1) a time stamp is included in the first column of the **.txy* files, and (2) the x and y coordinate values are not dependent on the previous coordinates; in other words, their values are absolute rather than relative as in the **.xyd* files. The advantage of the **.xyd* files over the **.txy* files is that the file size is cut in half. For our purposes, the **.txy* files are better because they provide the current x and y coordinate values.

Time	ULx	ULy	LLx	LLy	T1x	T1y	T2x	T2y	T3x	T3y	T4x	T4y	MNIx	MNIy	MNMx	MNMy
0	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000
6866	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000
13732	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000
20598	1000000	1000000	12162	-9741	-7800	5134	-27506	11521	-42724	9130	-54317	824	-654	-7450	-34804	-9839
27464	15757	11596	12175	-9772	-7786	5158	-27477	11512	-42758	9180	-54294	829	-699	-7452	-34792	-9862
34330	15760	11596	12186	-9803	-7776	5176	-27456	11503	-42796	9229	-54278	836	-741	-7453	-34782	-9883
41196	15763	11595	12187	-9835	-7770	5190	-27439	11494	-42833	9267	-54268	847	-781	-7454	-34772	-9904
48062	15766	11594	12181	-9871	-7765	5197	-27424	11481	-42864	9295	-54258	859	-818	-7455	-34765	-9925
54928	15768	11594	12170	-9903	-7758	5201	-27410	11468	-42885	9315	-54251	877	-854	-7456	-34761	-9943
61794	15771	11594	12158	-9933	-7753	5200	-27398	11455	-42898	9328	-54247	900	-886	-7457	-34760	-9959
68660	15774	11592	12146	-9954	-7752	5194	-27391	11441	-42905	9337	-54245	926	-916	-7459	-34764	-9972
75526	15776	11592	12133	-9966	-7761	5185	-27390	11431	-42910	9343	-54249	959	-944	-7459	-34774	-9981
82392	15779	11591	12124	-9969	-7777	5171	-27393	11425	-42914	9340	-54252	995	-969	-7461	-34787	-9988
89258	15782	11590	12117	-9967	-7803	5155	-27399	11430	-42918	9329	-54254	1032	-991	-7461	-34804	-9991
96124	15784	11589	12112	-9967	-7834	5135	-27407	11441	-42920	9309	-54259	1065	-1012	-7463	-34824	-9994
102990	15786	11588	12107	-9972	-7865	5117	-27411	11456	-42921	9284	-54269	1092	-1029	-7465	-34845	-9997
								⋮								

Figure 3.9: Excerpt from Task File *JW45_013.TXY*

Figure 3.9 shows an excerpt of a task file *JW45_013.TXY*, which describes subject JW45 performing task number 13. The headers for the columns are shown as the first row separated by a box; these headers do not exist in the actual file. The first column contains the time stamp (in 10^{-3}ms) for each row of pellet positions. The other columns contain the x and y coordinates of the pellets in the order:

```
ULx, ULy, LLx, LLy, T1x, T1y, T2x, T2y, T3x, T3y, T4x, T4y, MNIx, MNIy, MNMx, MNMy
```

where UL is the upper lip pellet, LL is the lower lip pellet, T1 to T4 are the tongue pellets (1 is closest to the tip and 4 is farthest back in the mouth), MNI is the front of the lower jaw (mandible incisor), and MNM is the back of the lower jaw (mandible molar). The pellet coordinates in the **.txy* files are given at a scale of 10^{-3}mm. The first few rows (as demonstrated in Figure 3.9) have pellet coordinates that represent a "bad data value" (1000000). This value is normally used when pellets are mistracked or when pellets do not exist, such as when pellets came loose during a recording session or were removed because their trajectories overlapped those of adjacent pellets. In the case of the "bad data values" at the beginnings of the X-ray pellet coordinate files, this was due to the sampling control software which governed pellet tracking and was not able to start at time zero.

Of the three additional files in the subject's directory, only two have currently been used: *Mis.dat* and *Pal.dat*. *Mis.dat* contains a log of all of the mistracked pellets. It is used as a guideline for selecting error-free files to be used as input to CASSI. *Pal.dat* contains coordinate values for the palatal outline which is drawn using our software. These files are shown in Figures 3.10 and 3.11.

Figure 3.10 shows Subject JW12's *mis.dat* file which provides a list of pellets which were mistracked and when they were mistracked within a specific record. The first column indicates the task number (or Record ID Number) where the error occurred. The second column indicates the pellet that was mistracked. Codes starting with 'T' indicate tongue pellets, the code 'UL' indicates the upper lip, the code 'LL' indicates the lower lip, codes starting with 'MAX' indicate reference pellets, and codes starting with 'MAN' indicate jaw pellets. The third and fourth column indicate the onset and offset of the mistracking, where -1 indicates the end of the file. Some rows have a fourth column. This is used for "raster hops", which occur when the XRMB loses one pellet because it has started tracking another pellet in its place. The *mis.dat* file was used as a guideline to determine which files were error-free or, in other words, which files had all of their pellets correctly tracked. This judgement was based on which task numbers did not occur in the first column of the *mis.dat* file. However, pellets that were removed were not included in *mis.dat*. Sometimes this led to the erroneous assumption that certain groups of files were "error free". Such was the case for subjects JW29 and JW32. Subject JW29 seems to have had the mandible molar (back jaw) pellet removed after the first five records. For this subject there seems to be only one usable file; thus, JW29 was not included in our tests. JW32 seems to also have had the mandible molar pellet removed after the first 19 or so tasks; we do, however, have four usable files for this subject.

The file, *pal.dat*, containing the palatal outline is shown in Figure 3.11. This file contains a sequence of x and y coordinates defining the palatal outline for subject JW12. These coordinates, which may vary in number from subject to subject, were obtained by at least one of two methods. One method involved taking a scan of the stone model (plaster dental cast) of the maxillary (upper) dental arch with a string of gold pellets laid along the mid-line of the palatal vault. From the scan, the pellet center locations were visually identified and the palatal curve was approximated by a piece-wise continuous function. The second method involved the experimenter or subject moving a tracing pellet along the palate mid-line while this pellet was being tracked by the XRMB system. The front-most and back-most points of the palatal outline were determined by the first and last pellets on the palatal chain for the stone model method and by the points where the speaker stopped tracing forward or back for the tracing pellet method.

010	TB1	0	767	
011	TB2	0	2366	
011	TD	14304	14889	
011	TD	2330	-1	k;j
012	TB2	0	-1	j;k
013	TB1	0	810	
013	TB2	0	-1	j;k
013	TD	0	5107	
015	TB1	0	1779	
015	TB2	0	1500	
015	TD	0	1957	
018	TD	0	1273	
018	TD	1253	-1	k;j
027	TB2	3294	3407	
028	TB2	1230	1420	
038	TB2	0	2792	
038	TD	0	1850	
038	TD	2771	-1	k;j
039	TB2	7900	8096	
039	TD	11216	-1	
043	TD	6678	6864	
044	TT	0	455	
050	TB2	2654	2819	
050	TB2	2795	-1	j;k
053	TB2	1905	3045	
057	TD	13356	-1	
069	TD	0	-1	k;j
073	TB2	0	-1	j;k
076	TB1	3299	-1	i;g
080	MAX_G	0	-1	
081	TT	0	-1	h;i
083	TB1	0	399	
083	TB2	0	399	
083	TD	0	399	
083	TT	0	399	
104	MAN_M	845	1674	
105	LL	1074	-1	
105	MAN_I	0	-1	
105	MAN_M	1018	1224	
105	MAN_M	1502	1755	
105	MAN_M	2955	3236	
105	MAN_M	4958	5220	
105	MAN_M	5936	6244	
106	LL	0	-1	
106	TT	0	-1	
107	TT	0	-1	
112	TT	0	-1	
116	MAX_G	2335	2582	
116	TB2	0	-1	j;k

Figure 3.10: Mistracked Pellets File (*Mis.dat*) for Subject JW12

-5000	8500
-10000	11500
-15000	16800
-19130	19622
-23902	22420
-29085	22902
-36661	23393
-41097	23250
-45000	22650
-47185	22298
-51185	21793
-53100	21333

Figure 3.11: Palatal Outline File (*Pal.dat*) for Subject JW12

Chapter 4

The CASSI Software System

In this chapter, we describe the CASSI (Computer Animated Speech SImulator) software system, which creates facial animations corresponding to X-ray microbeam data. Each data file records a sequence of positions of the tongue, lips, and jaw of a human subject. The CASSI system utilizes Parke's model of a human head, augmented by a representation of the inside of the mouth, including the tongue. CASSI's output is a 3D facial animation that shows, in animated form, the movements of a subject's articulators while speaking and performing other tasks.

This chapter provides details on the various versions of the CASSI system. Section 4.1 describes CASSI 1.0, June 1998, specifically, the initialization, animation and jaw rotation components used to connect the XRMB data to our augmented version of Parke's model. Section 4.2 describes CASSI 2.0, June 1999, in particular, the modifications made to the lips, which included rounding the lips and determining unique lip thickness values. Section 4.3 describes CASSI 2.1, September 1999, in regards to the thickness adjustments that provided certain subjects with more appropriate upper and lower lip thickness values.

4.1 CASSI 1.0: Initialization, Animation, and Jaw Rotation

The CASSI software system was designed to combine Parke's computerized facial model with X-ray microbeam data. Parke's code [23] allowed manual adjustments to be made to the parameter values controlling the shape of the face but did not provide animation. Using Marriot's code for ideas, double buffering and swapping (using the graphics library) were added to Parke's implementation to allow for smooth transition between animated frames. Animation of the face itself was created through frame by frame adjustments made to the values of the parameters. The goal was to create a software interface that allowed the XRMB data to drive the animation. To meet this goal, the frame by frame adjustments to the parameter values were based on the line by line changes in the XRMB pellets' locations.

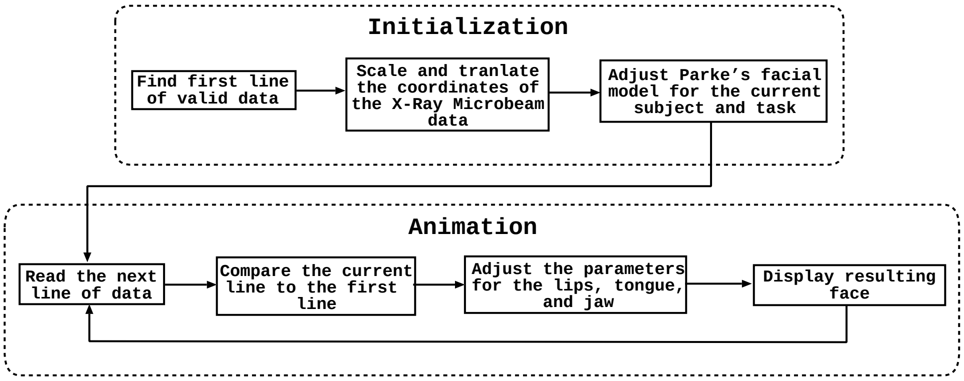


Figure 4.1: The Initialization and Animation Phases

Figure 4.1 shows the two major phases of execution in the CASSI software system: initialization and animation. These phases are shown in dotted boxes with the initialization phase on top; the animation phase on the bottom; and the interconnection between the two as an arrow from the last step in the initialization phase to the first step in the animation phase. The input to this system is one **.txy* file, which contains the pellet coordinate values for one speaker performing one task.

In the initialization phase, the vertices of Parke's facial model are positioned according to the individual subject and task. To make adjustments to the face, three steps are undertaken, as summarized in Figure 4.1. First, a row of valid data values is found in the **.txy* file; this step is required since each **.txy* file begins with conspicuously invalid data values indicating that the pellets have not yet been tracked. Once the first row of valid data values has been found, the second step scales and translates these data values to Parke's coordinate system. Finally, the third step adjusts the face. The shape of the chin and palate are adjusted based on the subject, and the initial positions of the lips, teeth, and tongue are adjusted based on the scaled and translated pellet locations.

Once the head's starting shape has been initialized and the lips, tongue, and jaw have been placed according to the pellets of the XRMB data, the animation can occur by making frame by frame changes to the starting face. The lower dotted box in Figure 4.1 shows the four major steps of the animation phase. The first step reads the next line of data (from the **.txy* file), containing the current x and y coordinates for the eight pellets. The second step computes the displacement of each of the current pellet positions from the initial pellet positions. In the third step, these displacements, which have been scaled, are used to adjust the parameters of the lips, tongue, and jaw. In the fourth step, the face is displayed with these adjusted parameter values.

Ideally, this animation phase is repeated for each line in the **.txy* data file. Depending on the speed of the graphical processing, however, some lines are ignored. On an SGI 02 computer with a 200MHZ R5000 CPU, every fifth line of data can be displayed. On a slower machine, namely a SGI Indigo with a 150MHZ R4400 CPU, every tenth line of data can be used.

In CASSI 1.0, the main focus is on the initialization and animation phases and on the jaw rotation procedure. The remainder of this section consists of four subsections. The first subsection discusses initialization, specifically: mapping the XRMB coordinate system to Parke's coordinate system; placing the lips, tongue, and teeth according to the pellet locations; drawing the palatal outline; and adjusting the chin for each subject. The second subsection discusses animation, in particular, the new parameters for the lips and tongue, and the parameters that have been disabled. The third subsection discusses the jaw rotation procedure for two reasons: (1) because it was one of the major focuses of version 1.0, and (2) because Parke's original jaw rotation function has been modified to move the chin and teeth in accordance with the motion of the jaw pellets. The last subsection briefly describes the preliminary tuning that was used to create this version.

4.1.1 Initialization

The major purpose of the initialization phase is to create each person's starting face. In the initialization phase, the XRMB coordinate system is mapped to Parke's model; the lips, tongue, and jaw are placed according to the pellets' starting locations; the palatal outline is drawn; and the chin and jaw are adjusted according to head measurements of the subject. These topics are now discussed in detail.

Mapping the XRMB Coordinate System to Parke's Model

Figure 4.2 shows the mapping used to translate the XRMB data to the 3D facial model, referred to as Parke's model since all of the components, except for the newly added tongue, were designed by Parke. The top, left snapshot in the figure represents the inside, side view of Parke's face including: the lips, part of the jaw, and the teeth (as polygons in the center of the diagram). The added tongue is seen as the dark line underneath the teeth. The x and z axes are denoted on the diagram. The z axis runs up and down and the x axis runs towards the front and back of the face. A white, two directional arrow indicates p, or Parke's jaw distance; this corresponds approximately to the distance between the two pellets on the jaw of the live subject.

On the left, directly below the snapshot of Parke's face in Figure 4.2, the XRMB data are shown. The x axis runs towards the front and back of the face, and y axis runs up and down. The palatal outline (or roof of the mouth) is shown as the dotted curving line, and the eight pellet locations are shown as open circles.

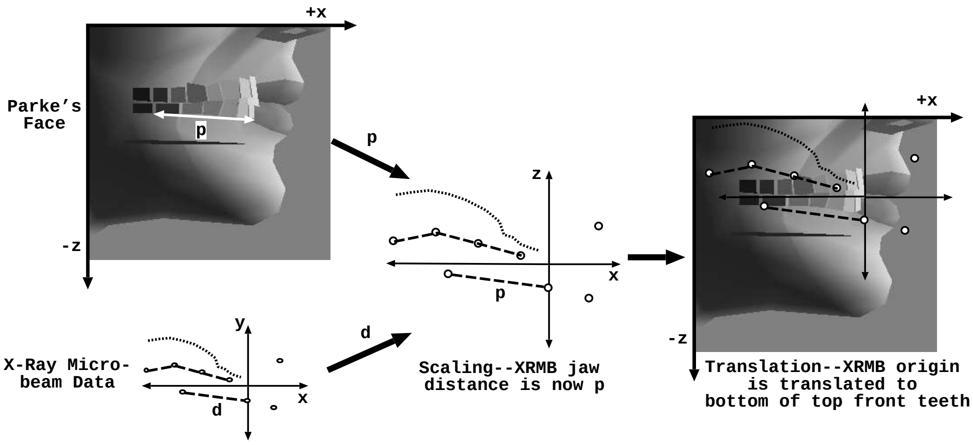


Figure 4.2: Mapping from XRMB Data to the 3D Facial Model

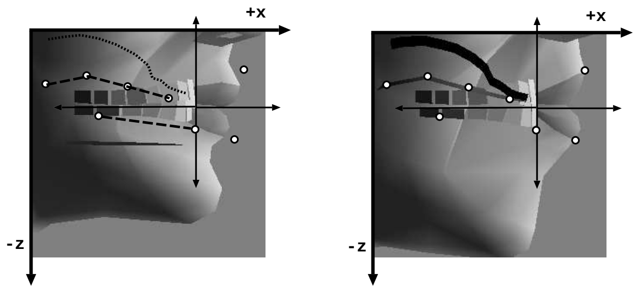


Figure 4.3: The Lips, Tongue and Teeth Adjusted According to the Pellet Positions

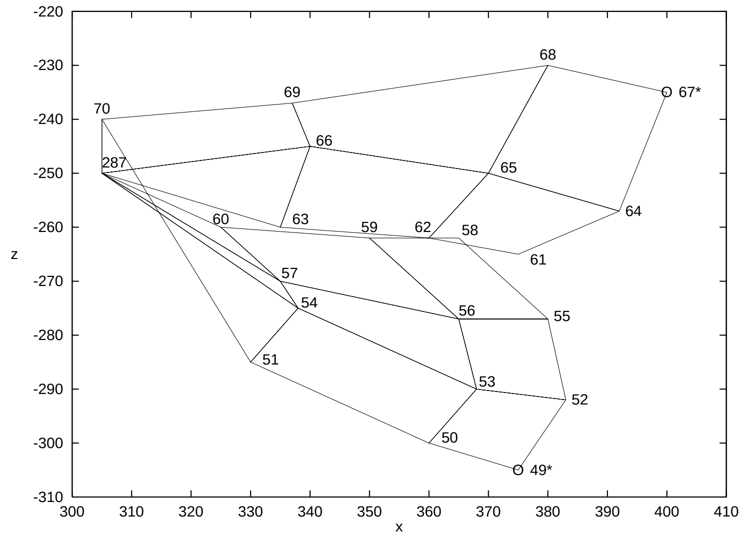


Figure 4.4: Side View of Lips

The eight pellets include: one for the upper lip and one for the lower lip (shown as the right-most open circles), two for the jaw (shown as the lowest connected open circles), and four along the tongue (shown as four points connected in a line directly beneath the palatal outline). The distance between the two jaw pellets according to the XRMB data is denoted by d.

Both coordinate systems, XRMB and Parke's, have x-axes that represent forward and backward motions. For the up and down motions, a correspondence is made between the XRMB y-axis and Parke's z-axis.

The center and right pictures in Figure 4.2 show the scaling and translation of the XRMB pellets. The scaling is based on multiplying each pellet's coordinate values by the ratio, p/d, where d is the distance between the two pellets on the jaw of the live subject, and p is the distance between two approximately corresponding points in the model. This scaling creates equal jaw lengths, p, between the facial model and the XRMB data, and it is assumed that all other pellets will be properly scaled. After scaling, translation is done so that the origin (for the pellet positions) is translated to the point corresponding to the bottom of the top front teeth in the facial model (vertex 226 as seen in Figure 4.7). The pellet positions now have a location relative to the 3D facial model.

Pellets and Their Associated Vertices

To initialize the face, the lips, teeth, and tongue are placed according to the scaled and translated pellet locations described in the previous section. Figure 4.3 demonstrates the placement of these components. The left side shows the scaled and translated pellet locations superimposed on Parke's original face. The right side shows Parke's face adjusted for subject JW40 so that the lips, teeth, and tongue correspond to the locations of these pellets. The right side also shows the palatal outline, discussed in the next section. To position the lips, teeth, and tongue, the pellets need to be associated with corresponding vertices of the facial topology. The following paragraphs describe the association made between the vertices and the pellets.

Figures 4.4 and 4.5 show the side and front views, respectively, of the lips inherited from Parke's topology. Each vertex is labelled with its associated vertex number. Two vertices, 67 and 49, are shown with open circles and their vertex numbers are marked with asterisks. These two vertices lie on the upper and lower edges of the lips where the pellets were placed. The upper lip (UL) pellet is associated with vertex 67 and the lower lip (LL) pellet is associated with vertex 49. The placement of these two vertices is adjusted to match exactly the scaled and translated UL and LL pellet locations. The other vertices of the lip are then adjusted accordingly. In addition, to compensate for high lips, the vertices corresponding to the tip of the nose are also adjusted to ensure that a space always exists between the upper lip and the nose.

For the CASSI system, Parke's facial model was augmented with 12 vertices representing the tongue. The top view of the tongue is shown in Figure 4.6. The tongue has been designed to fit inside the arch of

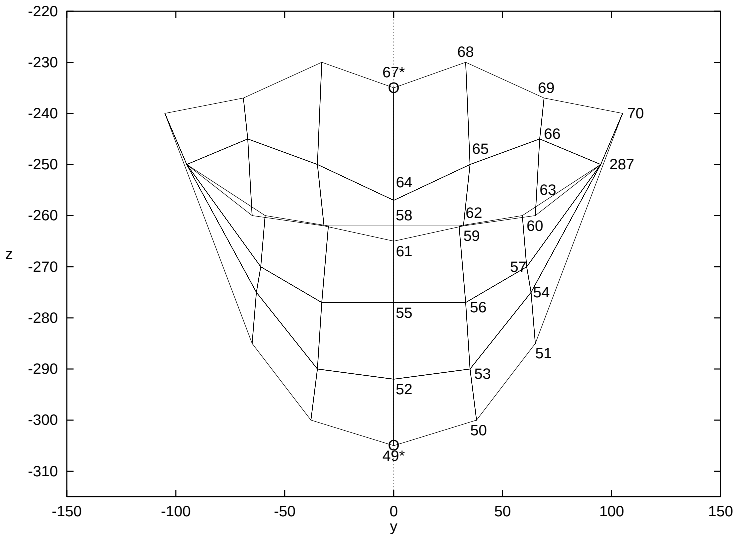


Figure 4.5: Front View of Lips

the bottom teeth (shown outside the tongue in Figure 4.6) and is flat, having no width in the z direction in Parke's coordinate system.

The tongue data for the XRMB system was dependent on four pellets attached along the central groove of each speaker's tongue. The most forward tongue pellet, T1, was placed behind the tip of the tongue. Accordingly, the tongue was designed with four vertices along the center to represent the pellets, and two additional vertices: one to represent the tip of the tongue, and the second to represent the back of the tongue. The tongue was then given width in the y direction by adding six additional vertices representing the outer edge of the tongue. The vertices corresponding to pellet locations are shown with open circles and asterisks beside their vertex numbers. Vertex 289 corresponds to T1, vertex 290 to T2, vertex 291 to T3, and vertex 292 to T4. These vertices are adjusted to match the location of the translated and scaled tongue pellet positions. The other vertices on the tongue are adjusted accordingly.

Figure 4.7 shows the side view of both the upper and lower set of teeth inherited from Parke's topology. The left side of the figure has polygons associated with the back teeth and the right side has polygons associated with the front teeth. Once again, the pellets mapped onto Parke's topology are shown by open circles and their vertex numbers are marked with asterisks. The first jaw pellet, MANi, was placed on the human subject on the outer surface of the central incisors where the gums and teeth meet. The MANi pellet is associated with vertex 258. The second jaw pellet, MANm, was placed on the human subject near the area between the first and second molars either where the teeth and gums meet, or on the gum itself. For simplicity, the pellet was assumed to be located where the teeth and gums meet. The MANm pellet is associated with a point halfway between vertex 286 and vertex 281. The modified jaw rotation function, as described in Section 4.1.3, is used to adjust the lower set of teeth.

Palatal Outline

During the initialization step, the palatal outline, or the roof of the mouth, is drawn using additional data from the *Pal.dat* file. Each subject has an individual palatal outline, described as a series of x and y coordinates in *Pal.dat* representing the palatal centerline. The number of x and y coordinate pairs varies among the subjects, because points were included in *Pal.dat* as needed to show the palatal outline. A three dimensional model of the palate was implemented, based on the two dimensions provided in the *Pal.dat* file and assumptions about the probable width and shape of the palate. The palate is narrow at the front of the mouth and wider toward the back. It has a slight downward slope from its centerline to its edges.

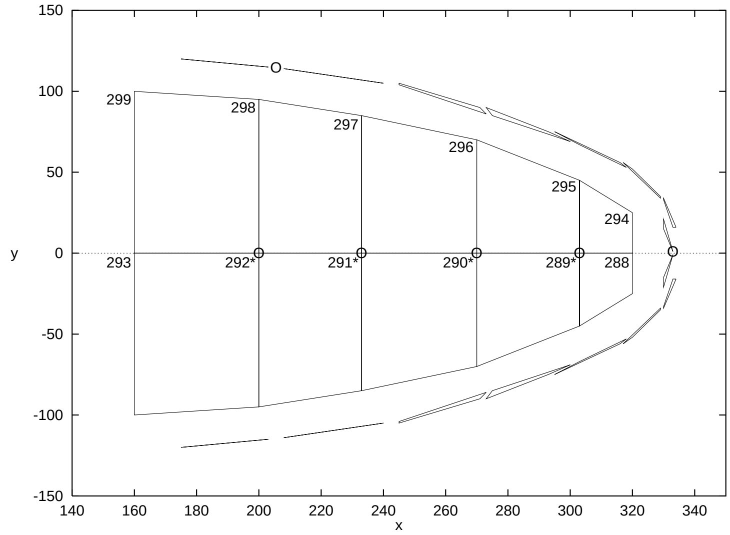


Figure 4.6: Top View of Teeth and Tongue

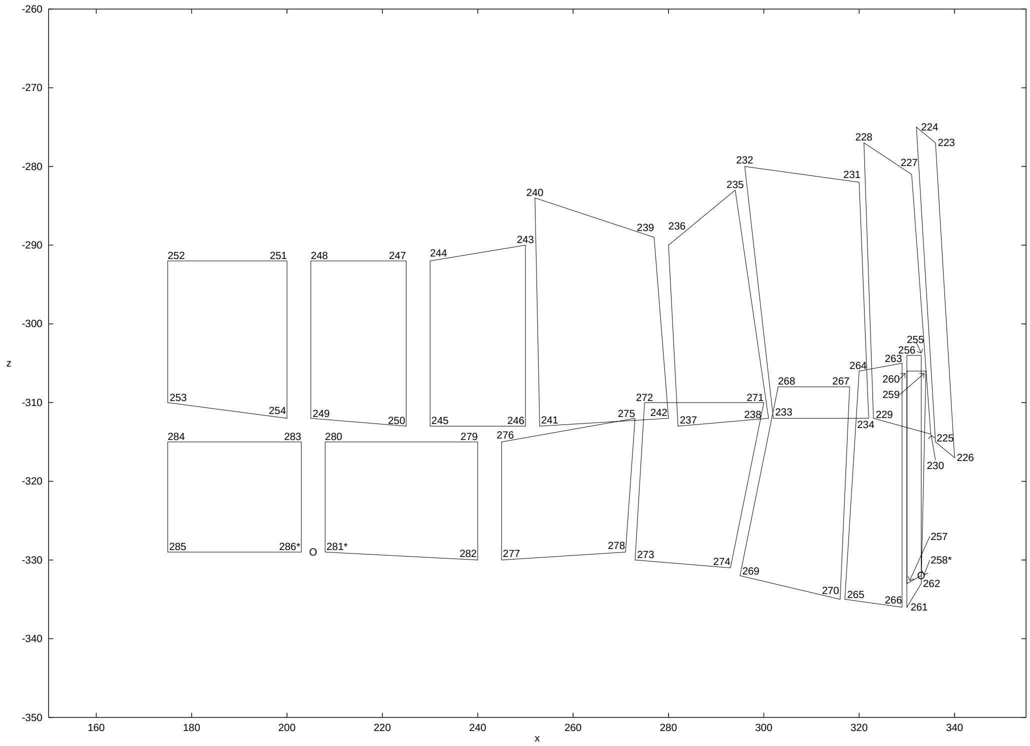


Figure 4.7: Side View of Teeth

Parameter #	Initial Value	Description
55	0.0	T1 X movement
56	0.0	T1 Z movement
57	0.0	T2 X movement
58	0.0	T2 Z movement
59	0.0	T3 X movement
60	0.0	T3 Z movement
61	0.0	T4 X movement
62	0.0	T4 Z movement
63	0.0	Upper Lip X movement
64	0.0	Upper Lip Z movement
65	0.0	Lower Lip X movement
66	0.0	Lower Lip Z movement

Table 4.1: Parameters Added to Parke's Original Set

Chin and Jaw Adjustments

Also, during the initialization step, the shapes of the jaw and chin are adjusted according to the measured positions of the gonion and gnathion, as given in the *Headmeasures.txt* file. Both the gonion and the gnathion are part of the jaw bone. The *gonion* is defined as the point at the angle of the jaw (in the back of the jaw, where the jaw angles upward towards the ears). The *gnathion* is the point of the chin. The shape of the jaw is adjusted using the jaw rotation equations described in Section 4.1.3.

4.1.2 Animation

Animation in CASSI is based on the movement of the XRMB pellets. Since each line in the XRMB **.txy* file contains the pellets' current locations, a new frame in the facial animation sequence could be generated for each line. To remain consistent with Parke's approach, these frames are generated through changes to the parameter values.

Although Parke's model had several parameters, these parameters were not designed to handle the XRMB data. The following shortcomings existed in Parke's set of parameters, with regards to input from the XRMB data: no parameters existed for the tongue pellets, no appropriate parameters existed for the lip pellets, a few of the original parameters interfered with the movement of the lips, teeth, and chin, and the existing parameter and procedure for jaw rotation did not handle XRMB data. These shortcomings led to three major changes: the addition of parameters for the lips and tongue movement; the adjustment to the functionality of the jaw rotation parameter and procedure; and the disabling of certain parameters that interfered with the movement of the lips, teeth, and chin.

Table 4.1 shows the parameters added for lip and tongue movement. The first column lists the parameter number, the second column lists its initial value, and the third column describes the parameter. In the third column, the tongue pellets are represented by T1, T2, T3, and T4, from front to back.

The functionality of the jaw rotation procedure has been adjusted. The original parameter for jaw rotation, parameter 4, still controls the jaw motion, but its value is now determined by the movements of the two jaw pellets. Instead of being fixed, the point of rotation for the jaw is now controlled by a new variable, *jawpoint*, which is calculated line by line from the **.txy* file. Most importantly, in jaw rotation, the lower lips are no longer rotated with the rest of the jaw. Jaw rotation is discussed in further detail in Section 4.1.3.

Some of the functionality of the original parameters had to be disabled so that they would not interfere with the newly added parameters and with the chin adjustments made during the initialization phase. These disabled parameters are summarized in Table 4.2. The first column lists the parameter number and the second column lists a brief description of this parameter. Parameters numbered 13, 14, 16, 17, 18, 21, and 22 were disabled because of their possible interference with the lips; parameters 47 and 48 were disabled because of their possible interference with the teeth; and parameters 19, 31, 32, 35, 36, and 46 were disabled because of their possible interference with the initialized chin.

The newly introduced parameters provide better control using the XRMB data than the original parameter set for three main reasons. First, parameters for the tongue are now included to allow the newly introduced tongue to move along with the XRMB data. Second, two parameters for the lower lip are included to make the lower lip motion independent of jaw rotation. Finally, the lip parameters are now controlled according to the data recording real lip motion; each lip (upper and lower) moves independently from the

Parameter #	Description
13	mouth interpolation
14	mouth X offset
16	mouth corner X offset
17	mouth corner Y offset
18	mouth corner Z offset
19	jaw Y scale
21	lower lip 'f' tuck
22	raise upper lip
31	chin X offset
32	chin Z offset
35	chin to mouth Z scaling
36	chin to eye Z scaling
46	growth factor
47	teeth X offset
48	teeth Z offset

Table 4.2: Original Parameters with Functionality Disabled

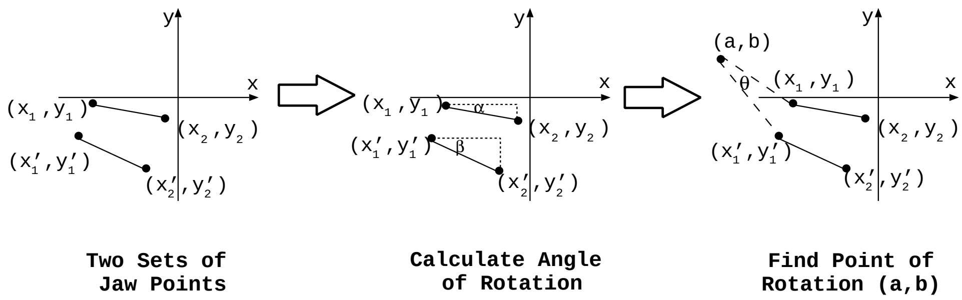


Figure 4.8: Calculating the Point of Rotation (a, b) for the Jaw

other, creating more varied motion than the previous set of parameters.

4.1.3 Jaw Rotation in CASSI 1.0

The underlying method for adjusting the jaw in both initialization and animation is the same. The method is based on the equations used by Parke [23] for jaw rotation as described in Section 3.1.3. In our implementation, these equations are used to derive a method of calculating the point of rotation, stored as *jawpoint*, for each frame of animation. The approach, which is shown in Figure 4.8, has two major steps: calculating the angle of rotation and finding the point of rotation. In Figure 4.8, (x_1, y_1) and (x_2, y_2) are the original jaw points and (x_1', y_1') and (x_2', y_2') are derived by rotating line $(x_1, y_1), (x_2, y_2)$ around a fixed point. To calculate the angle of rotation, the angle of incline of the original jaw position, α, and of the current jaw position, β, must be found. These angles are obtained using Equations 4.1 and 4.2:

$$tan\alpha = \frac{y_1 - y_2}{x_1 - x_2} \tag{4.1}$$

$$tan\beta = \frac{y_1' - y_2'}{x_1' - x_2'} \tag{4.2}$$

Using these two angles, the angle of rotation(θ) is determined:

$$\theta = \alpha - \beta \tag{4.3}$$

The next step is to find the point of rotation, (a,b). We begin with variations of Equations 3.3 and 3.4 used in Parke's model:

$$x' - a = (x - a)cos\theta + (y - b)sin\theta \tag{4.4}$$

$$y' - b = -(x - a)sin\theta + (y - b)cos\theta \tag{4.5}$$

From these, we derive the following equations for the point of rotation:

$$b = \frac{(y' + xsin\theta - ycos\theta)(1 - cos\theta) - (x' - xcos\theta - ysin\theta)(sin\theta)}{2 - 2cos\theta} \tag{4.6}$$

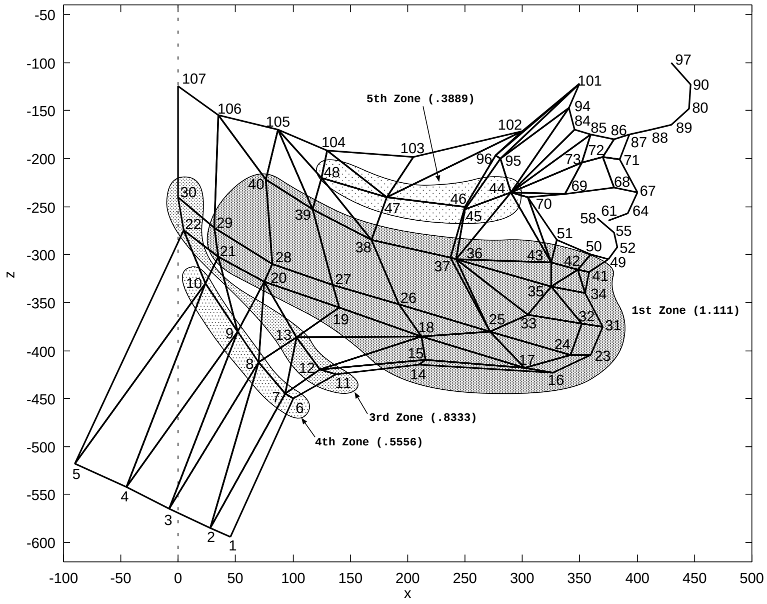


Figure 4.9: CASSI 1.0, Rotation Zones of the Cheek, Jaw and Neck (Side View)

Rotation Zone	Parke's Fraction of θ	CASSI 1.0 Fraction of θ	Vertices Rotated in CASSI 1.0	Description of vertices set
1st	1	1.11111	14 to 21, 23 to 29, and 31 to 43	jaw
2nd	.9	0.999999	255 to 286	lower set of teeth
3rd	.75	0.8333325	11, 12, 13 22, 30	underneath chin back of jaw
4th	.5	0.555555	6 to 10	neck vertices
5th	.35	0.3888885	44 to 48	lower cheek

Table 4.3: Summary of Vertices Rotated to Create Jaw Rotation in CASSI 1.0

$$a = \frac{y' + xsin\theta - ycos\theta - b(1 - cos\theta)}{sin\theta} \tag{4.7}$$

Once the angle of rotation and the point of rotation have been calculated, rotation zones similar to Parke's can be used to rotate the vertices of the teeth, jaw, cheeks, and neck. The rotation zones used by CASSI 1.0 are different from the rotation zones used by Parke in the following ways: different fractions of θ, corresponding to each zone, are used; vertices 22 and 30 have moved to the 3rd zone from the 1st zone; and the lower lip vertices are no longer rotated. In CASSI 1.0, Parke's fractions of θ are multiplied by 1.11111. This value is chosen to maintain the idea of Parke's rotation zones and to ensure that the teeth vertices, which correspond to the movement of the pellets placed on the jaw, have 100% jaw rotation, or a close approximation thereof.

These new rotation zones and fractions are summarized in Figure 4.9 and in Table 4.3. Figure 4.9 shows the side view of the chin, jaw, and neck and the sets of vertices belonging to the rotation zones. Figure 4.9 does not show the 2nd zone because it contains only the set of vertices for the lower set of teeth. Table 4.3 summarizes the set or sets of vertices in a rotation zone and the fraction of θ used for rotating. The first column contains the rotation zone label. The second and third columns have respectively the fraction of θ used in Parke's model and the Parke's fraction, multiplied by 1.11111, used in CASSI 1.0. The fourth column lists the set or sets of vertices belonging to the rotation zone, and the fifth column provides a brief description of the vertices.

The jaw rotation procedure handles both the initialization and animation phases for the jaw and teeth. A different choice of starting jaw location, (x_1, y_1),(x_2, y_2), and rotated jaw location, (x'_1, y'_1), (x'_2, y'_2), is selected for each of these two phases in order to determine the angle of rotation and point of rotation. In the initialization phase, the starting jaw and teeth locations are the original placement of the vertices in Parke's topology, and the rotated jaw and teeth locations are, respectively, from the placement of the jaw in *Headmeasures.txt* and from the location of the teeth, or jaw pellets, in the first line of valid X-ray microbeam data. In the animation phase, the starting jaw location is from the coordinates of the jaw pellets in the first valid line of X-ray microbeam data, and the rotated jaw location is from the coordinates of the jaw pellets in each consecutive line of data.

4.1.4 Preliminary Tuning

As described in the previous three subsections, the creation of CASSI 1.0 included the addition of vertices and parameters as well as a modified jaw rotation procedure. Preliminary tuning of the software was used to establish appropriate positions for the components of the tongue, lips, teeth, chin, neck and jaw.

Since jaw rotation was essential to the initialization and animation of four components, chin, neck, teeth, and jaw, most of the tunings in CASSI 1.0 focused on jaw rotation. Specifically, the exaggerated up and down jaw motion, or rotation, produced by task 106, the jaw wagging task, was tuned by examining the resulting front and side view animations created by this task.

Prior to the design of the current approach to jaw rotation, alternative approaches based on the rotation around a point representing the joint of the jaw were attempted. One approach used vertex 107, the original rotation point in Parke's model, as the fixed point of rotation. Another approach used two sets of jaw pellets from the XRMB data to dynamically calculate a point of rotation. To do this, a line equation was first derived from a set of jaw pellet positions, which consisted of the x and y coordinates for MANi and MANm. After calculating two of these jaw line equations, a point of intersection, which was assumed to be the joint of the jaw, was found. When compared to these other approaches, the current approach, as described in Section 4.1.3, produced more natural jaw movement.

Although little focus was placed on the lips, they were examined during animation to identify and ameliorate perceived anomalies, such as shadows or unusually placed vertices. These adjustments were made ad hoc with little methodology.

The result of these tunings was CASSI 1.0. Some problems, which are described in Section 5.2, exist in this version. Subsequent versions were created to improve the most visible component, the lips.

4.2 CASSI 2.0: Lip Movement

Three shortcomings were noted in the lip movement of CASSI 1.0: the animation of the lips was limited to up and down motions, the thickness of the lips was the same across all speakers, and the animated lips intersected with the surface of the teeth. CASSI 2.0 was designed to fix these three problems. First, elliptical outlines were used to create lip rounding, specifically, to move from an exaggerated "ee" sound, or stretched lip position, to a pucker for a kiss, or rounded lip position. Second, to create thicker or thinner lips, a unique lip thickness value was calculated for each subject. Third, to prevent the lips from intersecting with the surface of the teeth, the forward motion of the lips was controlled by a parabolic track.

To add lip rounding and lip thickness to CASSI 2.0, some preprocessing was required to extract information from the X-ray microbeam data. This preprocessing along with the animation and initialization phases are shown in Figure 4.10. The input to each phase is shown on the left, the operations are shown in the middle, and output is shown on the right.

As mentioned above, the three phases of CASSI 2.0 are preprocessing, initialization, and animation, as shown in Figure 4.10. The preprocessing phase is executed once to determine three key c values and to create two intermediate files, *maxprotru.txt* and *lipthick.txt*. These three key c values and two files are then used as input to the initialization phase. Initialization occurs once for each task file, which is associated with one subject. These task files are denoted by **.txy*, which indicates the extension assigned to these files in the XRMB data. In the initialization phase, the various components of the face are initialized according to (1) the coordinates of the pellets in the first valid line of data in the **.txy* file, and (2) the current subject's profile in *Headmeasures.txt*, *Pal.dat*, *maxprotru.txt*, and *lipthick.txt*. The output of the initialization phase is a display of the initialized face. The animation phase occurs next and is repeated for each subsequent

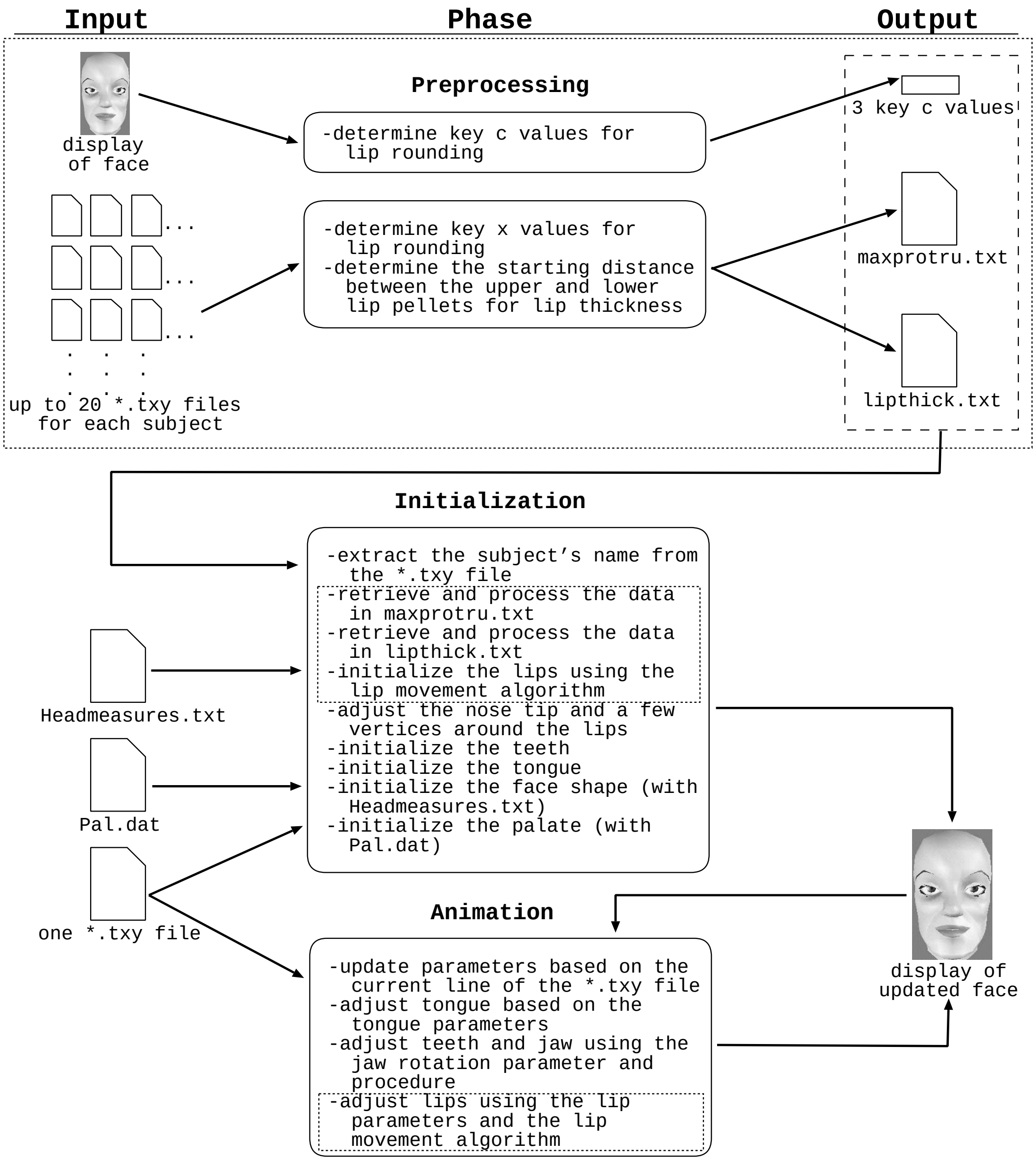


Figure 4.10: The Three Phases of CASSI 2.0

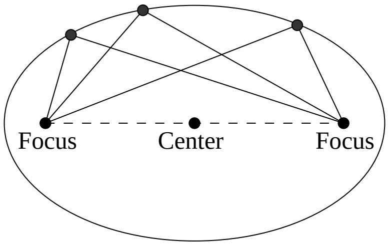


Figure 4.11: The Ellipse

line of the **.txy file*. In this phase, the next line of pellet coordinate values is read, and the parameters are adjusted. Through these parameters, the placement of the lips, teeth, tongue, and jaw are adjusted, and then the updated face is displayed.

The new components added to CASSI 1.0 to create CASSI 2.0 are denoted by a finely dotted box in Figure 4.10. These components include the preprocessing phase and the following steps in the initialization and animation phase: retrieving the key x values stored in *maxprotru.txt*, retrieving the lip thickness data stored in *lipthick.txt*, and initializing and adjusting the lips using the lip movement algorithm.

The lip movement algorithm is essential to CASSI 2.0. It contains the lip rounding, parabolic track, and lip thickness components used to determine the shape of the lips. Sections 4.2.1, 4.2.2, and 4.2.3 describe, respectively, lip rounding, parabolic track, and lip thickness. These sections are organized to describe the following: the mathmatical basis, if relevant; the preprocessing required, in the case of lip rounding and lip thickness; and the details of implementating these components. Section 4.2.4 provides an overview of the lip movement algorithm that combines these three components. Lastly, Section 4.2.5 describes the tuning used to create CASSI 2.0.

4.2.1 Lip Rounding

Waters, in his muscle-based model, used an elliptical shape to represent the muscles encircling the lips [28]. Assuming that Water's idea is valid and that the lip shape can be represented by ellipses, we control the shape of the lips through ellipses. The size and shape of the ellipse are determined by the current XRMB lip pellet coordinates and depend on three assumptions. The first assumption is that the lips are relaxed, or resting, at the first line of valid data in the XRMB **.txy* file. The second assumption is that when the lips move forward from the resting location, the lips round and the corners of the lips move towards each other. The third assumption is that when the lips move back, the lips stretch as the corners move away from each other.

The overall shape of the front view of the lips is controlled by an ellipse. In the following sections, the mathematical basis of the ellipse is described, the preprocessing used to extract lip information from the XRMB data is defined, and the method used to create both horizontal and vertical ellipses from the movements of the XRMB lip pellets are discussed.

Ellipse

An ellipse, as demonstrated in Figure 4.11, is defined by two fixed points called *foci*. It is formed by the points on a plane whose summed distance from each fixed focus is a constant. To visualize an ellipse, one can imagine a piece of string with the two ends tacked to the foci. A pencil is held tight against the string and is used to trace a curve. Since the sum of the distances to the foci is a constant, determined by the length of the string, the resulting curve will be an ellipse. If the foci coincide at the center, the ellipse becomes a circle.

To further define an ellipse, a, b, and c are used. Figure 4.12 shows these components, where c is the distance from the center of the ellipse to one of the foci, and $2c$ is the distance between the foci; and a and b depend on the major and minor axes respectively. The *major axis* is a line segment through the foci and across the ellipse; this is the longer of the two axes. The *minor axis* is a line segment across the ellipse, through the center, and perpendicular to the major axis. The length of the major axis is $2a$ and the length

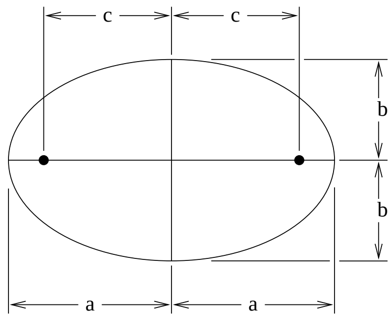


Figure 4.12: The Ellipse in Relation to *a, b, and c*

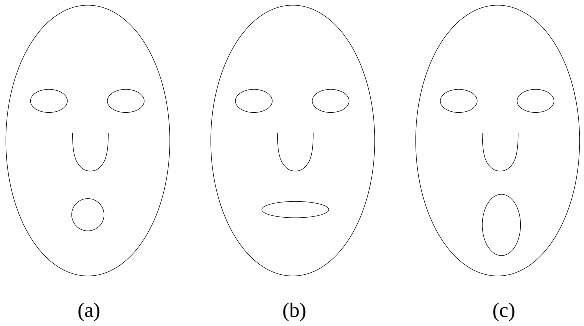


Figure 4.13: Three Major Lip Shapes

of the minor axis is $2b$. Equation 4.8 [2] provides a relationship between a, b, and c.

$$a = \sqrt{b^2 + c^2} \tag{4.8}$$

Equation 4.8 implies that $a \geq b$, where the equality holds true when $c = 0$. This equality occurs when the foci coincide.

We assume that the front view of the lips can be modelled with an ellipse. Large values of c create stretched lips, as occur when the vowel sound "ee" as in "bee" is made. Small values of c create rounded lips, as occur when the lips are puckered in a kiss. Figure 4.13(a) represents the rounded lips, Figure 4.13(b) the stretched lips. These two lip shapes can be represented by ellipses with their major axis parallel to the x-axis; we call these ellipses *horizontal ellipses*. Sometimes, *vertical ellipses*, or ellipses with their major axis parallel to the y-axis, are needed. Figure 4.13 (c) shows the vertical ellipse that may occur when the mouth is wide open.

Figure 4.14 shows the a, b, and c components of the horizontal ellipse on the left and the vertical ellipse

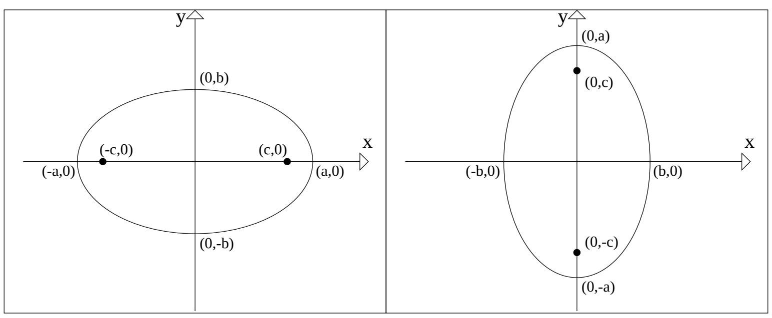


Figure 4.14: Two Kinds of Ellipse

on the right. The equations for these two ellipses are as follows [2]:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \tag{4.9}$$

$$\frac{x^2}{b^2} + \frac{y^2}{a^2} = 1 \tag{4.10}$$

Equation 4.9 corresponds to a horizontal ellipse shown on the left in Figure 4.14 and Equation 4.10 to the vertical ellipse shown on the right in Figure 4.14. Both of these equations are for ellipses with centers at the origin.

Sometimes, an ellipse has a center other than the origin. For an ellipse with its center at (h, k), we have the following equations [2]:

$$\frac{(x-h)^2}{a^2} + \frac{(y-k)^2}{b^2} = 1 \tag{4.11}$$

$$\frac{(x-h)^2}{b^2} + \frac{(y-k)^2}{a^2} = 1 \tag{4.12}$$

Equation 4.11 corresponds to a horizontal ellipse and Equation 4.12 corresponds to a vertical ellipse.

Preprocessing for Lip Rounding

The X-ray microbeam data describes 2-dimensional up and down, and forward and back motion. The third dimension, or the left and right movement as viewed from the front, must be estimated. To estimate this third dimension, the following previously mentioned assumptions are made: (1) the lips are in resting position at the start of the **.txy* file, (2) the lips round as they are moved forward from resting position, and (3) the lips stretch as they are moved backward from resting position. These rounded and stretched shaped lips are modelled by using a horizontal ellipse with varying values of c, where zero values produce round ellipses and large values produce stretched ellipses.

To create the varying lip shapes, we used key values, c_{rest}, c_{min}, and c_{max}, for each subject to represent lips that were resting, completely rounded, or stretched to their maximum. Key values, $c_{rest} = 65$, $c_{min} = 0$, and $c_{max} = 85$, were selected by viewing the computer animated face and determining positions which seemed natural for resting, rounded, and stretched lips.

For the lips to round according to changes in the data, the x positions of the XRMB lip pellets must be mapped to the corresponding c values. Three key x values from the XRMB lip pellets needed to be determined to correspond to c_{rest}, c_{min}, and c_{max}. For relaxed lips, the average starting location, x_{rest} is associated with c_{rest} based on the assumption that the lips are in a relaxed position at the start of any X-ray microbeam **.txy* data file. For rounded lips, the maximum x value, x_{max}, is associated with c_{min} based on the assumption that when the lips are most forward, they will be rounded. Finally, for stretched lips, the minimum x value, x_{min}, is associated with c_{max} based on the assumption that when the lips are pulled back, they will be stretched.

This correspondence between key c and key x values is summarized in Figure 4.15. In the first row, three extreme lip shapes, relaxed, rounded, and stretched are shown as ellipses. The second and third rows contain respectively, the key c and key x values for these three shapes as described above.

These three key values for x for each speaker were extracted from up to 20 error-free **.txy* files for that speaker. A file was considered to have errors if (1) it was listed in the mistracking file, *mis.dat*, or (2) it was found to have missing pellet data during processing. The remaining files were selected arbitrarily. The selected files are summarized in Table B.1 in Appendix B. If the file for the maximum lip protrusion task (task 118) was error free, it was included because it was likely to contain the maximum x value. In some cases, such as for JW32, a smaller sample was used because 20 error-free files did not exist.

As described in Appendix B, information about the starting, maximum, and minimum x-positions of the upper and lower lips was extracted and stored in a file called *maxprotru.txt*, which is shown in Figure 4.16. Each row of this file represents one subject. The first column of *maxprotru.txt* contains the subject's name. The second and third columns, entitled "REST_UL" and "REST_LL", represent the x locations for the resting upper lip and lower lip, respectively. The columns entitled "MAX_UL" and "MAX_LL" represent, respectively, the maximum upper lip and lower lip locations, corresponding to rounded lips. Since the lips will

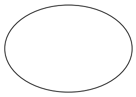
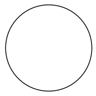
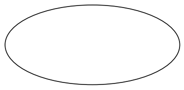
lip shape			
key c value (default)	C_{rest} (65)	C_{min} (0)	C_{max} (85)
key x value (basis in XRMB data)	X_{rest} (REST_UL)	X_{max} (MAX_UL)	X_{min} (MIN_UL)

Figure 4.15: The Correspondence between c and X-Ray Microbeam Data.

Subject	REST_UL	REST_LL	MAX_UL	MAX_LL	MIN_UL	MIN_LL
JW11	0.345274	0.228690	0.519758	0.439836	0.009925	0.003348
JW12	0.541010	0.390205	0.730061	0.485915	0.011575	-0.102688
JW15	0.452623	0.426632	0.606981	0.524212	0.008840	0.007371
JW16	0.431169	0.399125	0.600965	0.572621	0.008847	0.008228
JW18	0.536166	0.490561	0.834940	0.896159	0.427359	0.244083
JW19	0.492486	0.471077	0.684463	0.774152	0.371975	0.172287
JW21	0.335453	0.278352	0.502530	0.440367	0.008934	0.007164
JW24	0.365504	0.409801	0.621404	0.600709	0.007542	0.008639
JW25	0.363746	0.404857	0.668489	0.787208	0.007766	0.008383
JW27	0.349867	0.342236	0.639942	0.535901	0.007401	0.005606
JW32	0.445339	0.441625	0.621980	0.609274	0.284189	0.259924
JW40	0.528038	0.384745	0.767303	0.524264	0.011705	0.007449
JW41	0.477164	0.465303	0.821385	0.649093	0.406428	0.212140
JW45	0.462957	0.355955	0.692150	0.564495	0.010773	0.008101
JW502	0.319526	0.268207	0.582271	0.563677	0.264849	-0.037349

Figure 4.16: File *maxprotru.txt*

not always be maximally rounded, or brought forward, for each task, we assume that the best representative will be the greatest x value found in all the sample task files. Lastly, in Figure 4.16, the columns entitled "MIN_UL" and "MIN_LL" in the *maxprotru.txt* file represent respectively the minimum upper lip and lower lip locations, corresponding to stretched lips. Analogously with the "MAX" columns, we assume that the best representative will be the smallest x value found in all the sample task files. Of the data columns in *maxprotru.txt*, we use those related to the upper lip, namely, "REST_UL", "MAX_UL", and "MIN_UL" to determine x_{rest}, x_{max}, and x_{min} as described below. We currently do not use the lower lip data, but they are left in the file for completeness, and for possible future work.

Connecting the XRMB Data to the Ellipse

The previous section discussed the preprocessing phase, which selects the three key c values, and creates the file *maxprotru.txt*. In the initialization phase, the x values for a current subject are retrieved from *maxprotru.txt*. For both the initialization and animation phases, an equation that uses the correspondence between the key x values and key c values is used to estimate the roundness of the lips for the current line of the **.txy* file. This estimate, c_0, helps identify the ellipse shape used for the animated lips, and is typically adjusted by the algorithms that determine the a, b, and c values, which precisely define the elliptical outlines of the lips. Using the elliptical outlines, the points of the lips are adjusted, and the facial frame can then be displayed. The following section describes the retrieval and processing of the key x values, the conversion a current line of the **.txy* file into a c_0 value, the determination of a, b, and c, which define an ellipse, and the application of the elliptical outlines of the lips.

In the initialization phase, the key x values for a subject are retrieved from the *maxprotru.txt* file, and then are processed to fit into Parke's coordinate system. Specifically, the values in the row corresponding to the current subject are extracted and saved into the variables: $Rest\_UL$, Max_UL, and Min_UL. Then these values, which are relative to the XRMB coordinate system, are modified to fit Parke's coordinate system by scaling them according to Parke's jaw distance and translating them to the top front tooth, vertex 226, in Parke's topology. The result is three key x values, x_{rest}, x_{max}, and x_{min}, which are then placed in one to one correspondence with c_{rest}, c_{min}, and c_{max}.

Using this one to one correspondence and given the placement of the lip vertices 67 and 49 due to current line of the **.txy* file, an estimate of the roundness of the lips, c_0, can be found. Specifically, the current x value, x_{curr}, of the upper lip vertex, vertex 67, is converted into a c_0 value using the following equation:

$$c_0 = \begin{cases} \frac{(x_{curr}-x_{rest})(c_{min}-c_{rest})}{x_{max}-x_{rest}} + c_{rest} & \text{if } x_{curr} > x_{rest} \\ \\ \frac{(x_{curr}-x_{rest})(c_{max}-c_{rest})}{x_{min}-x_{rest}} + c_{rest} & \text{otherwise} \end{cases} \tag{4.13}$$

This equation involves translating and scaling from the x to c values.

To determine the shape of the ellipse, a, b, and c are calculated from Algorithm OUTER_ABC_CALC, shown in Figure 4.18. The height of the ellipse, stored as b, is calculated as half the vertical distance between vertex 67 and 49. The estimate of the roundness of the ellipse, c_0 is calculated using Equation 4.13. From this, a and c are calculated. Horizontal ellipses are easily created using c_0 value as c. However, vertical ellipses, which occur when the jaw is wide open, add some complications. As demonstrated in Figure 4.14, the c value is a vertical measurement for the vertical ellipse, rather than a horizontal measurement. When a horizontal ellipse becomes a vertical ellipse, "ellipse switching" occurs; this switch happens when the height of the ellipse is greater than the width. Ellipse switching is handled through the following equation, which calculates c using a variation of Equation 4.8:

$$c = \begin{cases} \sqrt{a^2 - b^2} & \text{if } b < a \text{ (i.e. height < width)} \\ \\ \sqrt{b^2 - a^2} & \text{otherwise (ellipse switch)} \end{cases} \tag{4.14}$$

In our notation, a always refers to the width and b always refers to the height of the ellipse, which varies from the notation used in Figure 4.14. This approach gives a, b, and c values which define the outer outline of the lips.

Given the outer outline of the lips, two other elliptical outlines are created to describe the frontview of the lips. The ellipse defining the outer edge of the lip is referred to as the "outer outline", the ellipse defining

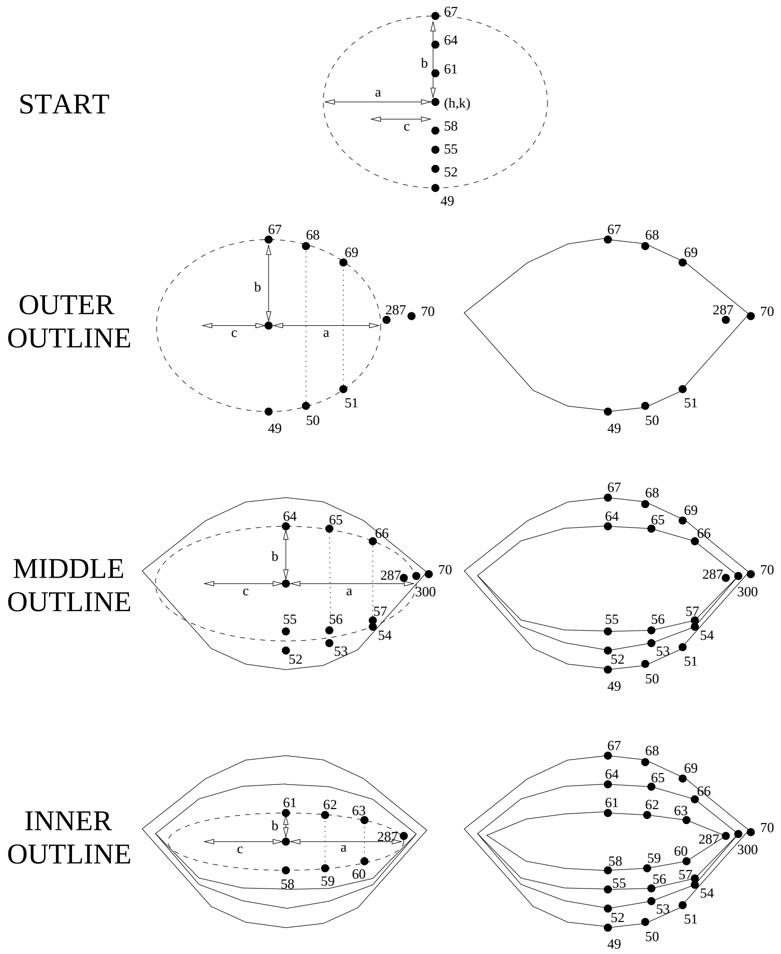


Figure 4.17: Using Ellipses to Model the Lips

Algorithm OUTER_ABC_CALC

Input: *preva*, *prevb*, *prevc* /* previous outer loop values */
vertices 67 and 49 /* upper and lower lip points */
x_{rest}, x_{max}, x_{min} /* from file *maxprotru.txt* */
c_{rest}, c_{min}, c_{max} /* used for calculating c_0 */
New_Subject /* flag for new task file */
Δ /* rate of change for c */
Output: outer outline's a, b, and c

1. calculate the height of the ellipse (stored as b); this is half the vertical distance between vertex 67 (ULy) and vertex 49 (LLy)
2. calculate c_0 using Equation 4.13
3. if *New_Subject*, then $c := c_0$ and a is calculated using Equation 4.8.
   else:
   (a) if the height of the ellipse is increasing and the c_0 is near c_{rest}, then assume that the lips are dropping (with little side to side movement of the corners). Do the following:
      i. decrease the width, $a := preva + ((prevb - b)/10.0)$
      ii. determine c using Equation 4.14:
   (b) else if the height of the ellipse is decreasing and the c_0 is near c_{rest}, then assume that the lips are rising (with little side to side movement of the corners). Do the following:
      i. increase the width, $a := preva + ((prevb - b)/10.0)$
      ii. determine c using Equation 4.14
   (c) else (the lips may be rounding or stretching along with moving up or down):
      i. set $c := \Delta c_0 + (1 - \Delta)prevc$
      ii. if the ellipse is horizontal (if $b < preva$, i.e height < previous width) or if c_0 is small (for round lips), use Equation 4.8 to calculate a
      iii. else (the ellipse may be vertical). In that case, set a and c:
         A. set $a := preva - ((prevb - b)/10.0)$
         B. determine c using Equation 4.14
4. set $preva := a$, $prevb := b$, and $prevc := c$

Figure 4.18: Algorithm OUTER_ABC_CALC

Algorithm MID_INNER_ABC_CALC

Input: $upper\_lip\_point$
$lower\_lip\_point$
$corner\_point$
Output: a, b, and c

1. calculate b:
$b := (upper\_lip\_point_z - lower\_lip\_point_z)/2.0$
2. calculate a:
$a := corner\_point_y$
3. calculate c using Equation 4.14

Figure 4.19: Algorithm MID_INNER_ABC_CALC

the vertical middle of the lip is referred to as the "middle outline", and the ellipse defining the inside outline of the lip is referred to as the "inner outline". Figure 4.17 shows these three outlines as dotted ellipses. The three major outlines are shown with left and right components: the left shows the dotted elliptical outlines, and the right shows the resulting outline of the lips.

The ellipse entitled "START" illustrates the vertices and variables that are required to subsequently generate the three ellipses. The required vertices and variables are as follows: vertices 67 and 49, which are positioned according to the current line of the XRMB **.txy* file; the center (h, k), which is determined by half the distance between vertices 67 and 49; the other lip vertices along the centerline, vertices 64, 61, 58, 55, and 52, which are distributed according to the locations of vertices 67 and 49 and the thickness of the lips; and the a, b, and c values for the outer outline, which are calculated using the algorithm OUTER_ABC_CALC, provided in Figure 4.18.

Algorithm MID_INNER_ABC_CALC, as shown in Figure 4.19, calculates the a, b, and c for the middle and inner outlines. The only required input to Algorithm MID_INNER_ABC is an upper lip point, a lower lip point, and a predetermined corner point. For the middle outline, the upper lip point is vertex 64, the lower lip point is a point midway between vertices 52 and 55, and the corner point is a new vertex, 300, that occurs midway between vertices 287 and 70. For the inner outline, the upper lip point is vertex 61, the lower lip point is vertex 58, and the corner point is vertex 287.

To summarize these two algorithms, the following can be noted: for the middle, inner and outer outlines, the b values are calculated as half the distance between an upper and a lower lip point; for the middle and inner outlines, the a values are determined by a corner point, and the c values are are determined exclusively by Equation 4.14; and for the outer outline, the a and c values are determined using Algorithm OUTER_ABC_CALC.

Using these three ellipses, defined by a, b, and c values, we can determine the front view of the lips, where y is the horizontal axis and z is the vertical axis. In general, to determine the (y,z) coordinates for the vertices, the y is calculated by evenly distributing the vertices between 0 and $2/3a$, and z is calculated using a variation of ellipse Equations 4.11 and 4.12. In our application, because b is the height and a is the width, these two equations are equivalent. The variation, written in terms of the original x and y, is:

$$y = \begin{cases} \sqrt{b^2 - \frac{x^2b^2}{a^2}} + k & \text{for upper lip} \\ -\sqrt{b^2 - \frac{x^2b^2}{a^2}} + k & \text{for lower lip} \end{cases} \tag{4.15}$$

For our application, the y becomes z and the x becomes y:

$$z = \begin{cases} \sqrt{b^2 - \frac{y^2b^2}{a^2}} + k & \text{for upper lip} \\ -\sqrt{b^2 - \frac{y^2b^2}{a^2}} + k & \text{for lower lip} \end{cases} \tag{4.16}$$

The following paragraphs describe, in further detail, how specific vertices are placed for each of the three outlines.

The outer edge of the lips, or the "outer outline" as described in Figure 4.17, has vertices 67, 68, 69, 70, 49, 50, and 51. The horizontal, y, coordinate values are first calculated by evenly distributing the vertices between 0 and $2/3a$ so that vertices 67 and 49 are at 0, vertices 68 and 50 are at $1/3a$, and vertices 69 and 51 are at $2/3a$. Vertex 70 is placed at a when the lips stabilize at the completely rounded position, i.e., when $c = 0$; otherwise, in order to stretch the lips and sharpen the corners, vertex 70 is placed increasingly far from a as c becomes larger. The vertical, z, coordinates, which create the outline of the ellipse, are then calculated by substituting the y values into Equation 4.16.

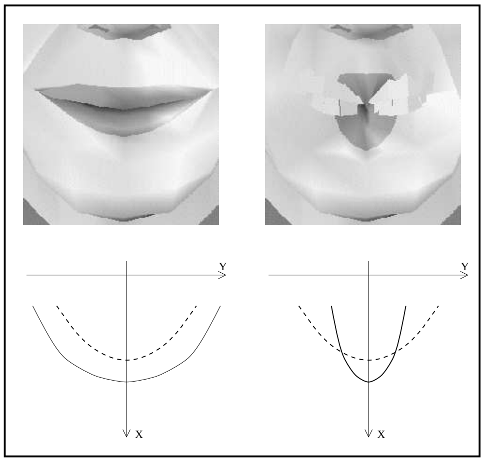
The center-line of the lips, or the "middle outline" as described in Figure 4.17, is defined by vertices 64, 65, 66, 55, 56, 57, 52, 53, and 54. The y value is based on an even distribution of the vertices between 0 and $2/3a$. For instance, vertices 64, 55, and 52 occur at 0; vertices 65, 56, and 53 occur at $1/3a$; and vertices 66, 57, and 54 occur at $2/3a$. The z values for the upper lip vertices can be determined by substituting the y values into Equation 4.16. The z values for the lower lip vertices are determined by using Equation 4.16 to find a point on the outline of the ellipse and then by placing pairs of vertices so that they are equal distance from this outline point. For instance, pairs 55 and 52, 56 and 53, and 57 and 54 are placed with equal distance on either side of the elliptical outline. This distance decreases as the pairs get closer to the corners. For example, the pair 57 and 54 is much closer to each other than pair 55 and 52.

Finally, the inside outline of the lips, or the "inner outline" as described in Figure 4.17, is defined by vertices 61, 62, 63, 58, 59, and 60. Once again the y values are based on the even distribution of the vertices between 0 and $2/3a$ so that vertices 61 and 58 occur at 0; vertices 62 and 59 occur at $1/3a$; and vertices 63 and 60 occur at $2/3a$. The z values are determined by substituting the y values into Equation 4.16. When lip vertices 58 and 61 cross each other at the centerline, or the lips overlap, Equation 4.16 is not used. In this case, the vertices on the inside outline are placed at equal distances from the centerline. For instance, if vertex 58 occurs above vertex 61 with a distance of 6, then vertices 61, 62, and 63 will be placed 3 below the centerline, and vertices 58, 59, and 60 will be placed 3 above the centerline.

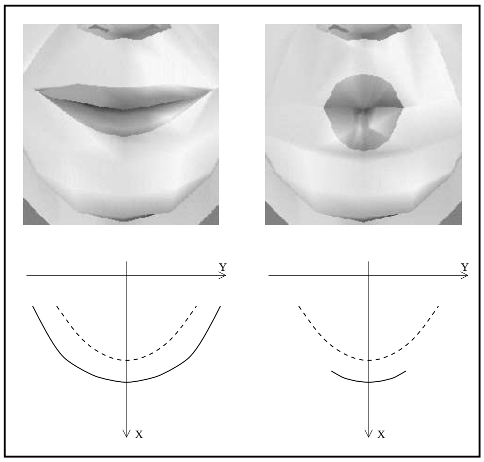
To create more natural movement, a few vertices around the lips are also adjusted. Above the lips, vertices 71, 72, and 73 are moved by portions of the movement of the upper lip vertices 67, 68, and 69, respectively. Below the lips, the x and z values of vertices 41, 42, and 43 are set according to half the influence of the jaw rotation algorithm and half the influence of the lower lip vertices 49, 50, and 51, respectively. The y values for these same vertices, 41, 42, and 43, are also influenced by the lower lip vertices 49, 50, and 51, respectively. Lastly, on either side of the lip, vertex 44, although influenced by other vertices, is most influenced by vertex 70, or the corner of the lip.

4.2.2 Parabolic Track

As described in Section 4.2.1, the y and z coordinate values for the front view are controlled by an ellipse. Some method of control is also required for the x, or forward, motion. Without any such method, the x values remain constant as the y and z values change. The result is that as the lips round, the corners move towards each other and "fold" together. Figure 4.20(a) demonstrates this "folding" lip movement, which causes the lips to poke through the surface of the teeth. On the left of Figure 4.20(a) is a snapshot of stretched lips, and on the right, rounded lips that have "folded" together. Below each snapshot, a corresponding graph is drawn. Each graph represents the top view of the lips, drawn as the solid curving line, and the teeth, drawn as the dashed curving line. The graph below the snapshot of the rounded lips in Figure 4.20(a) illustrates how the corner points are folded behind the surface of the teeth. To prevent the lips from "folding", a parabolic track is used, as demonstrated by Figure 4.20(b). The snapshot on the left of Figure 4.20(b) represents, once again, the stretched lips, and the snapshot on the right shows the rounded lips that no longer run into the surface of the teeth. The graph below the snapshot of the rounded lips in Figure 4.20(b) demonstrates that as the corners move closer, they also move forward on the "track", which prevents the lips from intersecting with the teeth. The following subsections describe this parabolic track method, in particular, the equations for the parabola, and applying these equations to the lips.



(a) without the Parabolic Track



(b) with the Parabolic Track

Figure 4.20: Lip Rounding

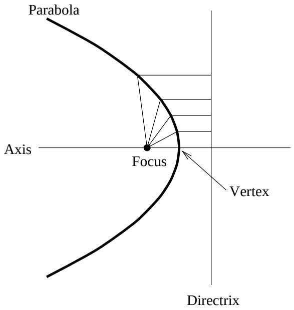


Figure 4.21: Features of the Parabola

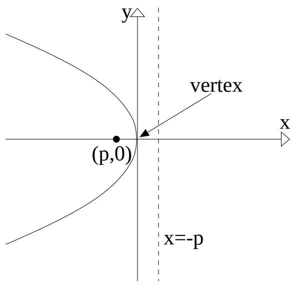


Figure 4.22: The Parabola with Vertex at the Origin

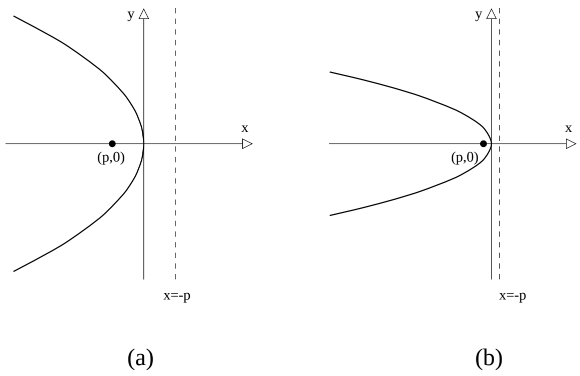


Figure 4.23: Two Parabolas with Different p Values

Parabola

A parabola is defined to be the set of all points in a plane that are equally distant from a given line and a given point not on the line [2]. Figure 4.21 shows the parabola with important features labelled. The given line is called the *directrix*, and the given point is called the *focus*. All points on the parabola are equally distant from the focus and directrix. The parabola is symmetric about the *axis*, or the line which passes through the focus at right angles to the directrix. The point where the axis meets the parabola is called the *vertex*.

A parabola may be in one of four different orientations: "U", "hill", "C" or "backwards C". The parabola that is used for our application is the "backwards C", as shown in Figure 4.22. For this particular parabola, the vertex is at the origin; the parabola opens in the negative x-direction and is symmetric about the x-axis; the focus is denoted by $(p, 0)$; and the directrix is denoted by $x = -p$. The value assigned to p is the distance between the focus and the vertex; in this case, p is negative. The equation corresponding to the parabola in Figure 4.22 is the following [2]:

$$y^2 = 4px \tag{4.17}$$

If the vertex, (h, k), is not at the origin, the corresponding equation is [2]:

$$(y - k)^2 = 4p(x - h) \tag{4.18}$$

As illustrated in Figure 4.23, the shape of the parabola is determined by the value of p. For instance, for a parabola with its vertex at the origin, values of p which are increasingly farther from zero, such as in Figure 4.23(a), yield wider parabolas than values of p which are closer to zero, such as in Figure 4.23(b). A p value that produces a parabola wide enough to represents the curve of the lips needs to be selected. For instance, the curve shown in Figure 4.23(b) is too thin to properly represent the curve of the lips because the lips poke through the surface of the teeth.

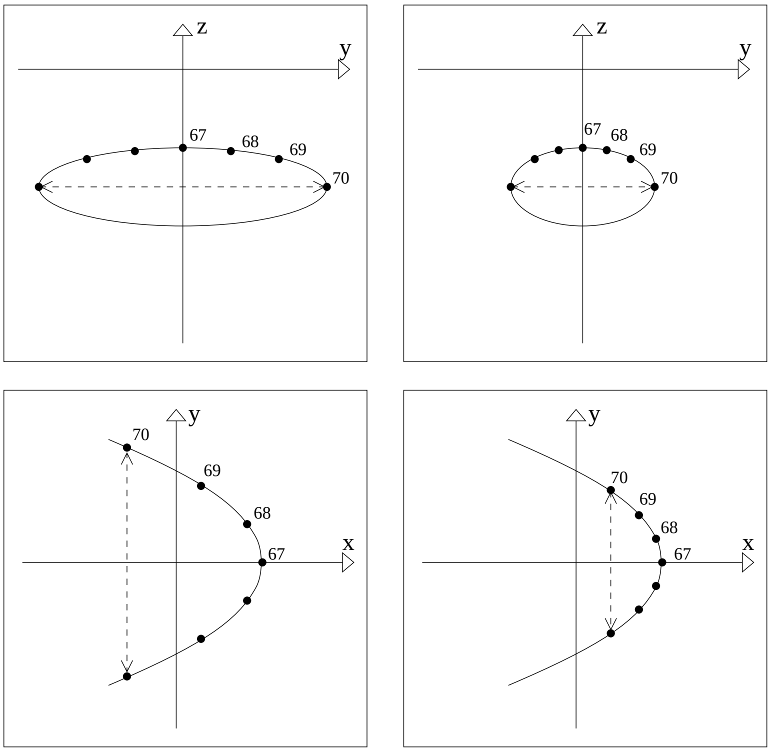


Figure 4.24: Applying the Parabolic Track to Stretched and "Rounded" Lips

Applied Parabolic Track

Figure 4.24 demonstrates the "parabolic track", which through its control of the x values, prevents the lips from poking through the surface of the teeth. The top two graphs in Figure 4.24 represent the outer outline of the lips from the front view, or the y and z axes. The bottom two graphs represent the parabolic track from the side view looking down on the lips, or the x and y axes. In all four graphs, lip vertices 67 through 70, which represent three upper lip points and one corner point, are labelled. The two graphs on the left illustrate the x, y and z coordinates when the lips are stretched, and the two graphs on the right illustrate the x, y, and z coordinates when the lips are more rounded. The dotted line represents the distance between the two corner vertices. On the left, this distance is large for the stretched lips, and the lip vertices, as shown on the bottom, have widely distributed x values. On the right, the distance between the corners is smaller because the lips are more rounded, and the lip vertices are moved forward. When the corners move closer to each other, the y distance is reduced, and the vertices slide forward on the "parabolic track". This prevents the vertices from intersecting with the teeth.

Given a y value determined from the ellipse calculations described in Section 4.2.1, an equation defining the parabolic track is used to find a corresponding x value. The following variation of the parabola Equation 4.18 is used to calculate x:

$$x = \frac{y^2}{4p} + h \tag{4.19}$$

where (h, k) occurs on the centerline of the lip. Thus, h depends on the x value of centerline vertices 67, 64, 61, 58, 55, 52, and 49, and k, which has been omitted in Equation 4.19, equals 0.

As previously mentioned, the value of p determines the shape of the ellipse. A satisfactory p value was chosen by systematically searching through values between 0 and -50 for a value closest to zero that prevented the lips from intersecting with the teeth, as observed by the experimenter. From this search, the chosen p value was -37, which created satisfactory results across all examined subjects.

	lip distance
JW27	0.442891
JW21	0.517419
JW25	0.414120
JW18	0.534753
JW12	0.629742
JW40	0.752903
JW502	0.491017
JW41	0.719188
JW15	0.579452
JW16	0.823341
JW11	0.975998
JW45	0.619474
JW32	0.503228
JW24	0.604045
JW19	0.785441

Figure 4.25: File *lipthick.txt*

4.2.3 Lip Thickness

With CASSI 1.0, a few subjects had upper and lower lips that never touched or lips that largely overlapped. To alleviate this in CASSI 2.0, a unique lip thickness was determined for each subject, and based on this thickness, the lips were adjusted by shrinking or expanding them in the z direction. Two assumptions were made in determining each subject's lip thickness. The first assumption was that the lips are lightly placed together in the first line of valid data in the XRMB **.txy* file. The second assumption was that the top and bottom lips have equal thicknesses. The following sections describe how the lip thickness was extracted from the XRMB data and how the lip vertices were placed using this thickness.

Preprocessing for Lip Thickness

A program was designed to extract a unique lip thickness value for each subject. This program relied on the same XRMB **.txy* files used for extracting the lip protrusion data and summarized in Table B.1. The program took as input all the **.txy* files for one subject and, from these files, determined the average distance between the upper lip pellet (UL) and the lower lip (LL) in the first line of valid data. The output of this program is the lip thickness value, or the the average distance from the top edge of the lip, UL, to the bottom edge, LL, for one subject. The program was run 15 times, once for each subject.

A compilation of the 15 lip thickness values was stored in a file called *lipthick.txt*, provided in Figure 4.25. The first column contains, in no particular order, the subject's name, and the second column contains the lip thickness value, normalized by the XRMB jaw length. This lip thickness value represents the average thickness of the upper and lower lips combined. Some subjects have thicker lips, such as JW11 with a lip thickness value of 0.975998, while others have thinner lips, such as JW25 with a lip thickness value of 0.414120.

Placement of the Lip Vertices Based on Thickness

Based on each subject's lip thickness value, the distances between the vertices of the lips are made to "stretch" or "shrink" along the z axis. This lip thickness value is extracted from the *lipthick.txt* file and undergoes two operations. First, the value is scaled from the XRMB coordinate system to Parke's coordinate system. Second, based on the assumption that the upper and lower lips are of equal thickness, the scaled lip thickness value, which represents the thickness of two lips, is divided by two. The resulting value depicts the thickness of one lip and is stored in the variable *Thick*.

Thick represents the thickness of one lip in relaxed starting position. From observation, when the lips round, they become thicker. To account for this, we have a local variable called *currthick*, which represents the current thickness of the lips. As the lips round, the value in *currthick* becomes larger than *Thick*. When the lips are not rounding, *currthick* and *Thick* are equal.

The distances between vertices along the centerline of the lips stretch or shrink in the z-direction to match the current thickness. Figure 4.26 demonstrates how the vertices move as the distances shrink. Figure 4.26(a) and (b) represent the side view of the lips and (c) represents the front view. All three views show the affected

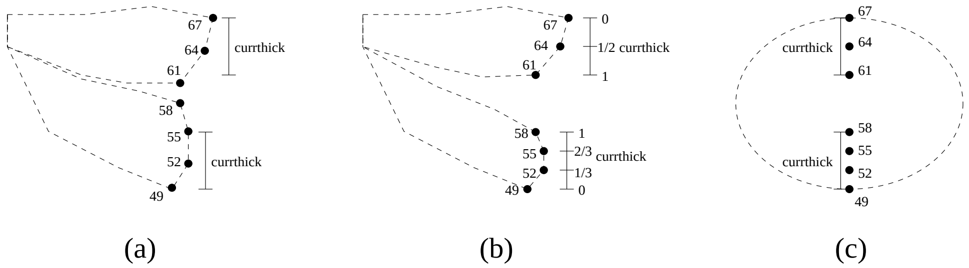


Figure 4.26: Adjusting the Lip Thickness

centerline vertices, 67, 64, 61, 58, 55, 52, and 49. Figure 4.26(a) shows vertices 67 and 49 as they have been positioned according to the upper and lower lip pellets, respectively; the other vertices have been placed relative to the positions of vertices 67 and 49 while maintaining the original topology's thickness. The current thickness, or *currthick*, is shown by two lines placed beside the lips: one has its edge aligned with vertex 67, and the other is aligned with vertex 49. Vertices 67 and 49 are not adjusted, but the other vertices on the centerline are adjusted according to the current thickness. Figure 4.26(b) shows the vertices from the side view, after they are adjusted to match the current thickness. The vertices are evenly distributed by portions of *currthick*; these portions are different between the upper and lower lips because the upper lip has three vertices whereas the bottom lip has four. For instance, for the upper lip, vertices 64 and 61 are positioned by half units of *currthick*; specifically, vertex 64 is positioned by $(vertex67 - \frac{1}{2}currthick)$ and vertex 61 by $(vertex67 - 1currthick)$. For the lower lip, vertices 52, 55, and 58 are positioned according to one third units of *currthick*; specifically, vertex 52 is positioned by $(vertex49 + \frac{1}{3}currthick)$, vertex 55 by $(vertex49 + \frac{2}{3}currthick)$, and vertex 58 by $(vertex49 + 1currthick)$. Figure 4.26(c) shows the newly positioned centerline vertices from the front view with the outer outline of the lips shown as a dotted ellipse.

4.2.4 Lip Movement Algorithm

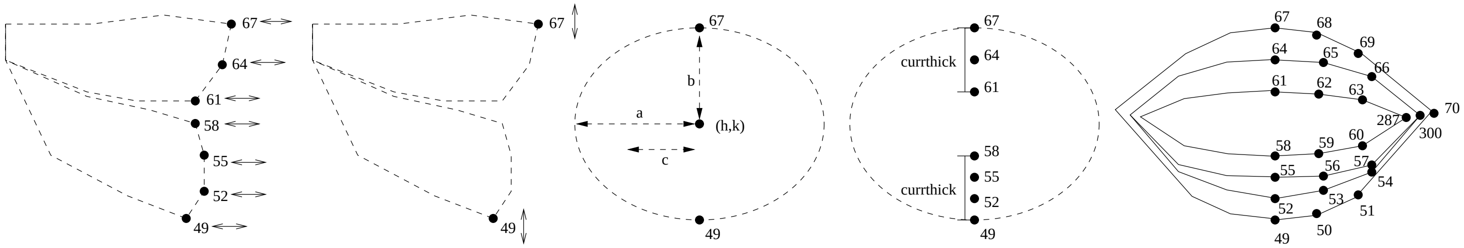
Figure 4.28 summarizes Algorithm LIP_MOVEMENT used in CASSI 2.0. This algorithm combines the elliptical outlines, parabolic track, and lip thickness components described in the previous sections. A corresponding diagram, shown in Figure 4.27, summarizes the algorithm in pictorial format. Figure 4.27(a), or steps 1 and 2, shows the lip vertices along the centerline moving forward or back. Figure 4.27(b), or steps 3 and 4, demonstrates the upper and lower lip vertices, 67 and 49, respectively, moving up or down. Figure 4.27(c), or steps 5 and 6, shows the "outer outline" with the corresponding center point, (h, k), and a, b, and c values identified. Figure 4.27(d), steps 9 and 10, shows the vertices along the centerline, 64, 61, 58, 55, and 52, adjusted according to the current thickness. Figure 4.27(e), or step 11, shows the front view of the lips after the y and z values have been determined using the elliptical outlines. Step 12 is not include in Figure 4.27 because it is difficult to illustrate the seven parabolic tracks, one for each centerline vertex, used to determine the x values of the lips.

4.2.5 Tuning

Because CASSI 2.0 focused on lip movement, the tuning for this version was meant to improve the motion of the lips. Particular attention was paid to the transitions between relaxed and rounded lips, normal and thickened lips, and horizontal and vertical ellipses. Adjustments were made to ensure that there was a smooth motion, without sudden changes, in the transition from one type to another.

For tuning, two task files were of particular interest: task 118, or maximum lip protrusion, and task 106, or jaw wagging. The maximum lip protrusion task, which created fully rounded lips at maximum protrusion, was used primarily to determine if the transitions between the relaxed and rounded lips and between the normal and thickened lips were smooth. The jaw wagging task was used to ensure that a smooth transition occurred when the shape of the lips switched from a horizontal to a vertical ellipse.

The tuning itself consisted of viewing a small set of task files, noting the anomalies, and making appropriate adjustments. When satisfactory animations were obtained on the small set of task files, more task



(a) Steps 1 & 2 (b) Steps 3 & 4 (c) Steps 5 & 6 (d) Steps 9 & 10 (e) Step 11

Figure 4.27: Overview of Lip Movement in CASSI 2.0

Algorithm LIP_MOVEMENT

Input: newpts[MAX_POINTS][3] /*copy of the input vertices adjusted to the subject */
pm[63], pm[64], pm[65], pm[66] /*parameters for the upper and lower lip movement*/
/*plus all input required for Algorithm OUTER_ABC_CALC*/

Output: pt[MAX_POINTS][3] /*set of vertices with lips adjusted for the current frame*/

1. move vertices 61, 64, and 67 forward or backward by pm[63], which corresponds to the x movement of the upper lip pellet
2. move vertices 49, 52, 55, and 58 forward or backward by pm[65], which corresponds to the x movement of the lower lip pellet
3. move vertex 67 up or down by pm[64], which corresponds to the y movement of the upper lip pellet
4. move vertex 49 up or down by pm[66], which corresponds to the y movement of the lower lip pellet
5. find the center, (h, k), of the ellipse, where h will always be 0 and k is the vertical midpoint between vertex 67 and vertex 49
6. calculate a, b, and c for the outer outline of the lips using Algorithm OUTER_ABC_CALC in Figure 4.18
7. place limits on the a and c values calculated in step 6 to restrict the shape of the lips
8. thicken the lips by adjusting $currthick$ to be greater than $Thick$ when the lips are rounding, or, specifically, when $c_0 < 50$
9. stretch or shrink the distance between the upper lip vertices, 67, 64 and 61, according to the current thickness, $currthick$
10. stretch or shrink the distance between the lower lip vertices, 49, 52, 55, and 58, according to $currthick$
11. calculate the front view of the lips, or the y and z values, using the three ellipse outlines summarized in Figure 4.17
12. calculate the x values by using the parabolic track demonstrated in Figure 4.24

Figure 4.28: Algorithm LIP_MOVEMENT

files were added to the set and then tested. First, a set of three maximum protrusion task files was tested. Next, two jaw wagging task files were added to the set. Then, files of arbitrarily selected other tasks were included. Lastly, when all these task files had satisfactory results, more task 118 and 106 files were added.

Often adjustments corrected one problem, while creating another. For testing, in these cases, the set of task files was first limited to those specific problem files. Once the problem was corrected, other task files were added to the set. Eventually, after many viewings and adjustments, all the task files used for testing produced satisfactory animations without major anomalies. CASSI 2.0 resulted from this tuning.

4.3 CASSI 2.1: Adjusted Lip Thickness

The major difference between CASSI 2.0 and 2.1 is the method of determining the thickness of the lips. In CASSI 2.0, the upper and lower lips are assumed to be of equal thickness. In CASSI 2.1, certain subjects whose lips were centered particularly low, are adjusted so that the upper lip is not as thick as the lower lip. The following sections describe the criteria used for adjusting the lips, the placement of the vertices based on the different thickness values, and the blending functions that were used to smooth the movement of the corners of the lips for those modified subjects.

4.3.1 Criteria for Adjustments

Some subjects were considered to have well centered lips while others were considered to have poorly centered lips. A subject whose lips were well centered had equal amounts of the upper and lower set of teeth showing most of the time, or, in other words, the lips were evenly centered between the gap made by the upper and lower teeth. When animated by CASSI 2.0, subjects who had poorly centered lips were clearly distinguished by lips that went below the bottom front teeth. The software was modified to adjust the thickness of the lips for these subjects.

Specifically, to determine which subject's lip thickness needs to be modified, a comparison is made between the placement of the inside edge of the upper lip and vertex 226, or the tip of the top front teeth. For this comparison, the upper lip is assumed to have an equal thickness to the lower lip, as it would have in CASSI 2.0. If the starting placement of the inside edge of the upper lip exceeds a threshold of 20 points below vertex 226, then the top lip is considered to be too big and is made smaller.

Some subjects did not have their upper lip below vertex 226. For these subjects, the lip thickness does not change from CASSI 2.0 to 2.1. The following sections are relevant only for those subjects whose lip thickness is adjusted.

4.3.2 Placement of the Vertices

For review, $Thick$ represents the thickness of one lip, and $currthick$ represents the current thickness, which is modified from $Thick$ when the lips round. Both these variables assume that the upper and lower lip are of equal thickness. The current combined thickness of the two lips, can, thus, be written as $2 * currthick$.

In CASSI 2.1, the upper and lower lips are not of equal thickness. The thickness of the upper lip, denoted by $currthick\_upp$, is one third the combined thickness of the lips, or $2/3 * currthick$. The thickness of the lower lip, denoted by $currthick\_low$, is two thirds the combined thickness of the two lips, or $4/3 * currthick$. These different thickness values are used to stretch or shrink the distances between the vertices of the lip.

Figure 4.29 illustrates the differences and similarities in the placement of the vertices from CASSI 2.0 to CASSI 2.1. All four pictures in the figure represent the side view of the upper and lower lips. The top two pictures, taken from Figure 4.25, represent CASSI 2.0, and the bottom two represent CASSI 2.1. The major difference between these two versions is the values of the upper and lower lip thicknesses. In CASSI 2.0, the upper and lower lip have equal thicknesses, $currthick$; whereas, in CASSI 2.1, the upper and lower lips have different thicknesses, $currthick\_upp$ and $currthick\_low$, respectively. All these thicknesses are shown in Figure 4.29 as lines placed beside the lips and aligned with vertex 49 and vertex 67. A similarity can be identified in the distribution of the vertices from CASSI 2.0 to 2.1. The distributions are illustrated as labelled fractions on the thickness lines in the two pictures on the right of Figure 4.29. For the upper lip in CASSI 2.0 and 2.1, the lip vertices are distributed by half units of the thickness value, or $currthick$ and $currthick\_upp$, respectively. For the lower lip, the lip vertices are distributed by one third units of the

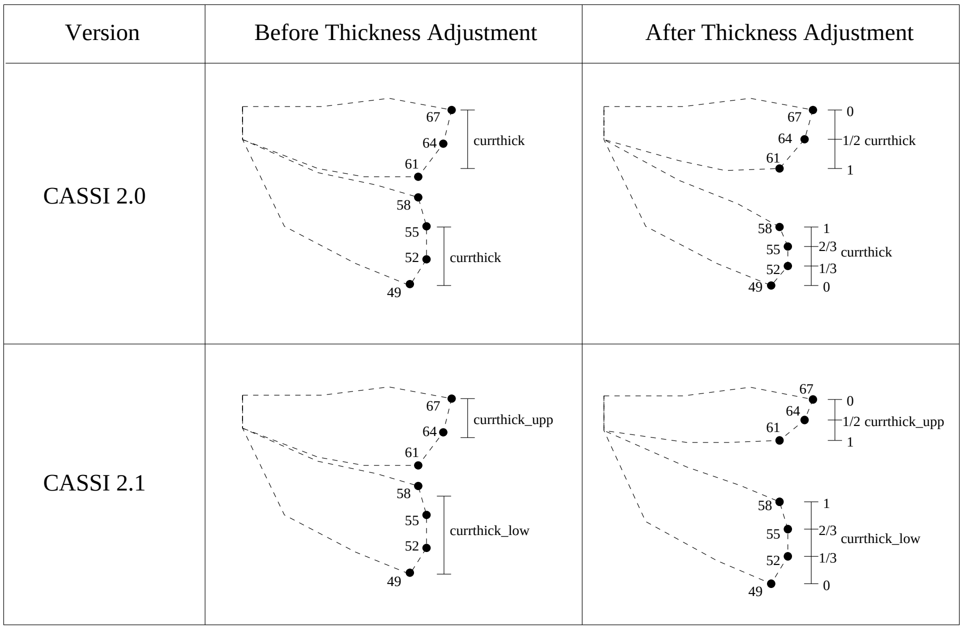
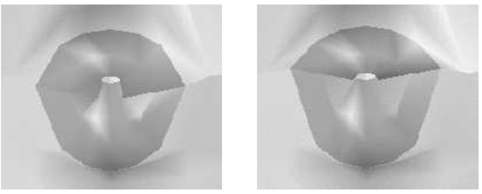


Figure 4.29: The Thickness Adjustments in CASSI 2.0 and 2.1



(a) **(b)**

Figure 4.30: JW40's Rounded Lips (a) with Poorly Placed Vertex 287 (b) with Blending Function Placement

thickness value, or *currthick* in CASSI 2.0 and *currthick_low* in CASSI 2.1. As seen by Figure 4.29, the idea behind placing the vertices remains the same; the difference lies in the thicknesses used for the upper and lower lips.

4.3.3 The Blending Functions

The lip thickness adjustments greatly improved the centering of the lips for some subjects. However, when the lips were fully rounded, these modified subjects had irregular shaped lips, as shown in Figure 4.30(a). This irregular shape was caused by the inside corner vertex, 287, passing below vertex 60, the vertex on the inside outline of the lower lip that is closest to the corner. To fix this problem, as the lips rounded, blending functions were used to influence the z value of vertex 287 by the z value assigned to vertex 60. Eventually, at rounded position, vertex 287's z-value was equivalent to that of vertex 60. The improvement in the lip shape made by using the blending functions is shown for comparison in Figure 4.30(b).

The blending functions used in CASSI 2.1 are shown in Figure 4.31. These functions were chosen because they are nonlinear and, thus, provide subtle changes, or smooth transitions, at the extremes, when $t = 0$ or 1. The function denoted by $F1$ in Figure 4.31 has decreasing values as t increases, and the function denoted by $F2$ has increasing values as t increases.

$F1$ and $F2$ are two of the four Hermite blending functions used to determine a cubic polynomial curve. The Hermite curve is constrained by two end points and two tangent vectors at the endpoints. Since CASSI 2.1 is only concerned with the influence of two endpoints, only the two blending functions related to the end

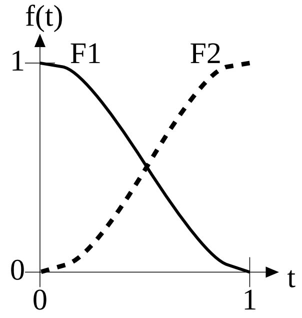


Figure 4.31: Blending Functions F1 and F2

points are required [13]:

$$F1 = 2t^3 - 3t^2 + 1 \tag{4.20}$$

$$F2 = -2t^3 + 3t^2 \tag{4.21}$$

The sum of these two functions is 1. Applied, these functions can be used as weights for determining the amount of influence of two components.

The inside corner vertex, 287, is influenced by two vertices: the original vertex, 287, and vertex 60. As the lips become rounded, the influence of vertex 287 decreases as the influence of vertex 60 increases. To create these decreasing and increasing influences, the original vertex 287 is weighted by the blending function F1, and vertex 60 by F2. The t value is based on the c_0 determined by Equation 4.13. Specifically, the c_0 values occur between 0 and 50. As an example, when the lips first round, c_0 is slightly less than 50, $t \approx 0$, vertex 287 has close to 100% influence, and vertex 60 has very little influence. Similarly, when the lips are fully rounded, $c_0 = 0$, $t = 1$, vertex 60 has 100% influence, and vertex 287 has no influence. The blending functions are, thus, used to gradually change the z-value of the inside corner vertex, 287, from the z-value of the original vertex 287 to that of vertex 60. The inside corner thereby moves smoothly upwards, which improves the shape of the lips at fully rounded position.

Chapter 5

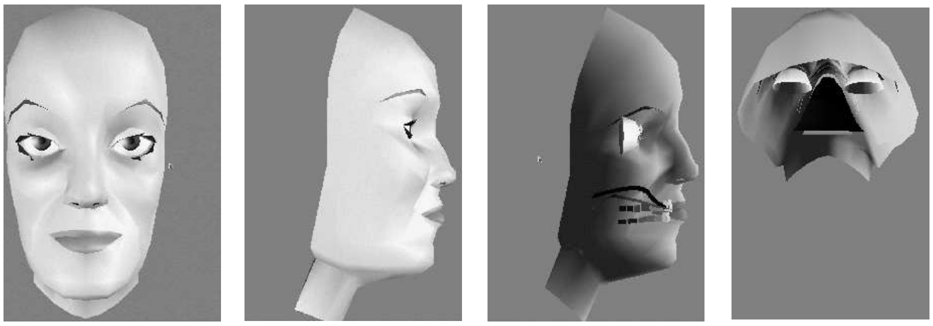
Results

Three versions of CASSI were created for this research. CASSI 1.0 focused on the initialization and animation of the lips, teeth, jaw, and tongue, with particular focus on jaw rotation. CASSI 2.0 improved the motion of the lips in the following ways: the lips were made to round, the lips were prevented from running into the surface of the teeth, and the lips were given a thickness that depended on the particular subject. For CASSI 2.0, the thickness of one lip was determined by half the lip thickness value. The final version, CASSI 2.1, improved the treatment of the lip thickness by making the upper lip one third of the lip thickness value and the lower lip two thirds of the lip thickness value for certain subjects.

These versions were subjectively evaluated to determine where improvements were needed and where improvements had been made. This chapter contains the method and results of evaluations performed on CASSI 1.0, 2.0, and 2.1. Section 5.1 summarizes the method of evaluation, the tasks used in the evaluation, and the rating system. Section 5.2 provides the description and results of the evaluation performed on the teeth, lips, and tongue for CASSI 1.0. This section also provides an evaluation of jaw and palate placement. Section 5.3 provides the evaluation of CASSI 2.0 and 2.1. This section includes a comparitive evaluation of CASSI 1.0, 2.0, and 2.1, an evaluation of the maximum lip protrusion task (task 118) and the jaw wagging task (task 106) across the three versions.

5.1 Evaluation Method

To evaluate the CASSI system, the facial animations for each subject were viewed from four positions. Each facial animation was viewed from the front (viewing the full face from the starting orientation, as in Figure 5.1(a)), the side (viewing the full face from a side orientation, as in Figure 5.1(b)), the inside (viewing



(a)Front View (b)Side View (c)Half Face from Side d)Top of Palate

Figure 5.1: Different Views of the Face Used in the Evaluation

Subject	Task #	Description
JW11	19	sentences
	21	citation words
JW12	2	citation words
	5	citation words
JW15	9	citation words
	10	sentences
JW16	3	number sequences
	6	citation words
JW18	4	citation words
	14	citation vowels
JW19	1	citation words
	23	citation words
JW21	7	sentences
	12	paragraph
JW24	17	sentences
	18	citation words
JW25	15	vowel sequences
	16	citation VCV
JW27	20	sentences
	22	citation words
JW32	13	citation sVd's eg. side
	15	vowel sequences
JW40	5	citation words
	8	citation words
JW41	1	citation words
	11	paragraph
JW45	3	number sequences
	13	citation sVd's eg. side
JW502	21	citation words
	22	citation words

Table 5.1: Subjects and Tasks Used for Evaluation

half of the face from a side orientation, as in Figure 5.1(c)), and the top (viewing the top of the palatal outline, as in Figure 5.1(d)). Appendix A describes how these positions were obtained using the interface. During the evaluation, parameters were not manually adjusted.

A subjective evaluation was performed to assess the quality of the facial animation. This evaluation made it possible to determine where improvements were needed, and (in the case of CASSI 2.0 and 2.1) where improvements had been made.

These evaluations were based on viewing the animations created by input from a **.txy* file, which corresponds to one subject performing one task. The **.txy* files were arbitrarly selected so that each of the 15 sample subjects had two error-free files corresponding to two different tasks. Table 5.1 shows the subjects and tasks used for evaluating CASSI across the different versions. The first column of the table lists the subject. The second and third columns provide the task number and a brief description of the task, respectively. For instance, the brief description of task 3 for subject JW16 is "number sequences"; specifically, this refers to the sequence "9739286 8495571 5945341" read as individual digits separated by pauses.

The evaluation focused on three components of the face: lips, tongue and teeth. These components were viewed from different orientations and their strengths (denoted by "+") and weaknesses (denoted by "-") were noted. The strengths were not used in the rating system, but they were used to comment on the strengths of a component of the face or the improvements made to a component. The rating system was based on the number of faults and was the following: EXCELLENT, VERY GOOD, GOOD, FAIR,

Rating	# of Negatives
EXCELLENT	0
VERY GOOD	1
GOOD	2
FAIR	3
POOR	4
VERY POOR	5
UNACCEPTABLE	$\geq$6

Table 5.2: Summary of the Rating System

POOR, VERY POOR, or UNACCEPTABLE. EXCELLENT means that there was nothing wrong with that component, VERY GOOD means that one problem was seen for that component (one –), GOOD means that two problems were seen (two –'s), FAIR means that three problems were seen (three –'s), POOR means that four problems were seen (four –'s), VERY POOR means that five problems were seen (five –'s), and UNACCEPTABLE means that six or more problems were seen (six or more –'s). This rating system is summarized in Table 5.2. Specific details about this rating system are discussed in Sections 5.2 and 5.3.

5.2 A Subjective Evaluation of CASSI 1.0

The major focus of CASSI 1.0 was the initialization and animation of the lips, teeth, jaw, and tongue. Thus, CASSI 1.0 was evaluated on the placement and movement of these components. Two evaluations, summarized by tables, were performed. First, in Section 5.2.1, the movement of the teeth, lips, and tongue for the tasks described in Table 5.1 is evaluated. Second, in Section 5.2.2, the overall appearance of the face, particularly, the initialized placement of the jaw and palate is evaluated. In addition, Section 5.2.3 provides a discussion of these results.

5.2.1 Teeth, Lip, and Tongue Evaluation

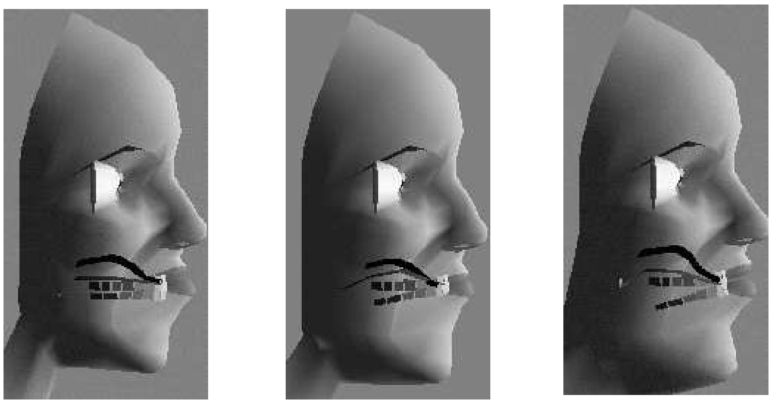
An evaluation, summarized by Table 5.3, was performed on the lips, teeth, and tongue for the subjects and tasks described in Table 5.1. Table 5.3 was constructed by examining the animations for anomalies, and by comparing among subjects. The first column lists the subject. The second, third, and fourth columns summarize, respectively, the positions and movements of the teeth, lips, and tongue. In the table, the lower set of teeth are evaluated by their position in relation to the upper set of teeth; the lips are evaluated by their initialized placement and by their movements; and the tongue is evaluated by its movement, in particular, whether or not it pokes through the surface of the teeth. The following pararagraphs present the details of the evaluation performed on the teeth, lips, and tongue.

For the evaluation of the teeth positions, the focus of the evaluation was the side view looking inside the mouth. From this view, close attention was paid to the positioning of the lower set of teeth while the task was being performed, and particularly at the end of the task. Subject JW24 was considered to have ideal positioning for the teeth because the bottom teeth were horizontally aligned with a very slight overlap in the front teeth (see Figure 5.2). All other subjects were compared to JW24. For instance, when compared to JW24, JW27 (in Figure 5.2) had overlapping front teeth (this was one –) and lower teeth that sloped upward from back to front (this was another –). These two factors (two –'s) caused JW27 to have an overall rating of GOOD for teeth position. Another example is JW41 who had a very steep slope upward from back to front, when compared to the other subjects (see Figure 5.2). Since this slope was extremely large, three negatives were assigned to it; this gave JW41 an overall rating of FAIR for teeth position.

For the evaluation of the lip positions, the front and side views were used. As shown in Figure 5.3, the front view (without adjustments in orientation) was examined to see if the lips lightly touched (desirable), were too far apart (as in JW11), or were too close together (as in JW502). The front view was also examined during the animation to determine if the lips were well centered (desirable) or if they were centered too high or too low (undesirable); and to determine if the lips passed through the surface of the teeth (undesirable). The centering (high, low, or well) was determined by the position of the lips when they were opened. If

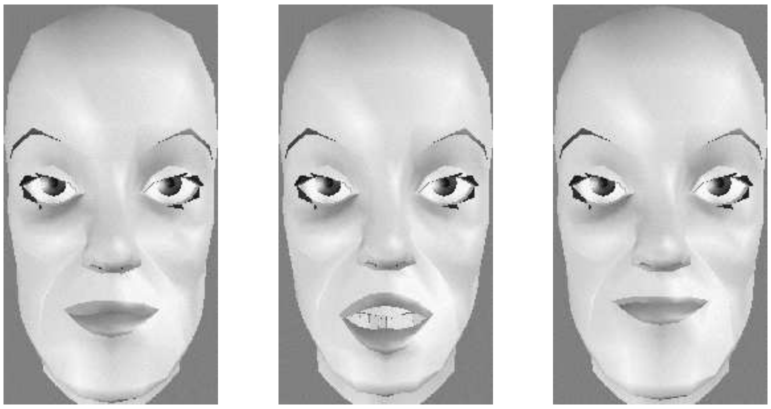
Subject	Teeth Position	Lip Position	Tongue Position
JW11	VERY GOOD –back teeth overlap (top to bottom)	VERY POOR –lips go through surface of teeth – –there is a large separation between lips (they never touch) – –lips are centered low (often go below the bottom of the lower front teeth)	FAIR – – –tongue pokes noticeably through upper and lower teeth
JW12	GOOD –back teeth overlap(top to bottom) –bottom teeth slant downwards towards front teeth	GOOD – –lips are centered low (go below the bottom of the lower front teeth)	VERY GOOD –in task 2, tongue pokes through upper front teeth
JW15	GOOD –front teeth overlap (top to bottom) –bottom teeth slant upward towards front teeth	GOOD –lips are separated at beginning –lips go below the bottom of the lower front teeth	GOOD –the tongue pokes through the upper front teeth (more noticeable in task 10) –from the front view, the tongue is barely viewable (hidden by teeth)
JW16	VERY GOOD –bottom teeth slant upward toward front teeth	FAIR – –there is a large separation between lips –lips are centered high	VERY GOOD –tongue pokes through top front teeth only in task 3
JW18	FAIR –front teeth overlap (top to bottom) – –bottom teeth slant upward toward front teeth	GOOD –lips are tightly placed together –lips are centered high	GOOD – –for task 14, tongue pokes slightly through upper and lower teeth
JW19	VERY GOOD –bottom teeth slant upward toward front teeth	GOOD – –there is a large separation between lips	GOOD – –tongue pokes through upper and lower teeth (more noticeable in task 23)
JW21	VERY GOOD –teeth overlap (top to bottom) +bottom teeth horizontally aligned	VERY GOOD –lips go through surface of teeth	FAIR – – –tongue pokes noticeably through upper and lower teeth
JW24	EXCELLENT +bottom teeth horizontally aligned	FAIR +from front view, lips are nicely closed (for task 18) – –from side view, lower lip protrudes farther than upper lip –lips are well centered but go below the bottom of the lower front teeth	FAIR – – –tongue pokes noticeably through upper and lower teeth
JW25	VERY GOOD –bottom teeth slant upward toward front teeth (slight)	FAIR –lips are tightly placed together –upper lip goes through the surface of the upper set of teeth –from side view, lower lip protrudes farther than upper lip	VERY GOOD –In task 16, tongue pokes through lower teeth
JW27	GOOD –front teeth overlap (top to bottom) –bottom teeth slant upward toward front teeth	VERY GOOD –lips are tightly placed together	GOOD – –tongue pokes through upper and lower teeth
JW32	VERY GOOD –teeth overlap (top to bottom) +bottom teeth horizontally aligned	FAIR –lips go through the surface of teeth –from side view, lower lip protrudes farther than upper lip (slight) –lips are tightly placed together at beginning	EXCELLENT
JW40	VERY GOOD –bottom and top teeth seem too far apart +teeth are evenly separated (top to bottom)	VERY GOOD –there is a separation between lips at beginning	VERY GOOD –in task 8, tongue pokes through upper teeth
JW41	FAIR – – –bottom teeth slant upward toward front teeth	FAIR –there is a separation between lips at beginning – –lips are centered low (go below the bottom of the lower front teeth)	GOOD – –tongue pokes through upper teeth for both tasks and the lower teeth in task 11
JW45	VERY GOOD –bottom teeth slant upward toward front teeth (slight)	EXCELLENT	GOOD –in task 13, the tongue pokes through the upper front teeth –from front view, tongue is barely viewable (seems too high above the top teeth)
JW502	VERY GOOD –bottom teeth slant upward toward front teeth (slight)	VERY GOOD –lips are tightly placed together	EXCELLENT

Table 5.3: Subjective Evaluation of Subjects and Tasks in CASSI 1.0



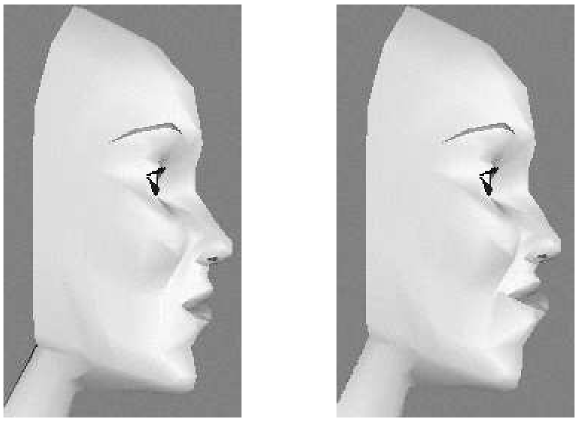
JW24 **JW27** **JW41**

Figure 5.2: Internal Side View of EXCELLENT (JW24), GOOD (JW27), and FAIR (JW41) Teeth Positioning



JW45 **JW11** **JW502**

Figure 5.3: Front View of EXCELLENT (JW45), FAIR (JW11), and VERY GOOD (JW502) Lip Positioning



JW45 **JW24**

Figure 5.4: Side View of EXCELLENT (JW45) and FAIR (JW24) Lip Positioning

JW12 **JW27** **JW24**

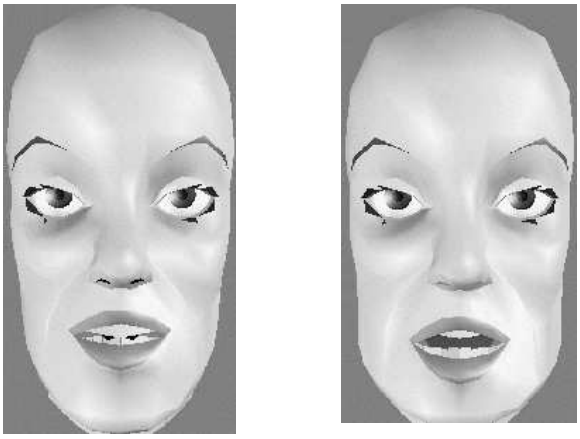
Figure 5.5: Front View of Tongue Poking through the Teeth for JW12 (VERY GOOD) JW27 (GOOD), and JW24 (FAIR)

the lips were centered on the gap between the teeth (or equal amounts of the upper and lower teeth were displayed) at most times, then the lips were considered to be well centered (desirable); if the lower lip often went below the lower set of teeth, then the lips were considered to be centered low; and if there was a large amount of the upper set of teeth showing, and little of the lower set of teeth showing, then the lips were considered to be centered high. The side view (see Figure 5.4) was examined to see if the bottom protruded more than the top (as in JW24). Again, for situations with significant problems, extra negatives were assigned. For instance, when compared to others, JW11 (in Figure 5.3) had lips that were widely separated and, in fact, never touched. For this large lip separation, JW11 was assigned two –'s. JW11 also had lips that went below the bottom of the lower front teeth (two –), and lips that went through the surface of the teeth (one –). This gave JW11 an overall rating of VERY POOR (with five –'s).

The tongue position was evaluated only from the front view. The front view was examined (from the original starting location) to see if the tongue poked through the teeth (see Figure 5.5). If the tongue poked through either the upper or the lower front teeth (as in JW12), then one negative was assigned. If the tongue poked through both the upper and lower front teeth (as in JW27), then two negatives were assigned. If the tongue poked very noticeably through the upper and lower front teeth (as in JW21), then three negatives were assigned. Also from the front view, one negative was assigned if the tongue was barely viewable (as in JW15). JW502 and JW32 were rated as EXCELLENT because, from the front view, the tongue was visible and did not appear to poke through the teeth.

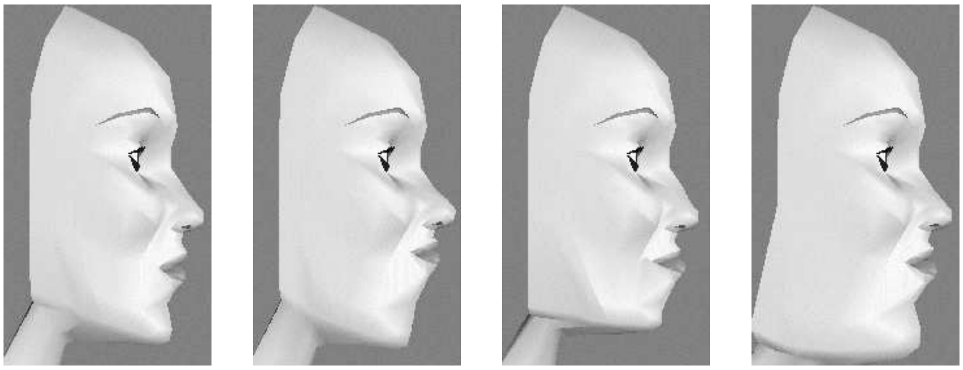
The tongue was not evaluated with regard to its shape from the internal side view or whether it poked through the palatal outline. From the internal side view, the tongues for all tasks for all subjects (with the exception of task 5 for JW40) had a V-shape in the back at times during the animation (this V-shape can be seen in subject JW41 in Figure 5.2). This V-shape results from the initialization of the farthest back points in the tongue. When the tongue is initialized, the farthest back points are "anchored" to a position that is farther back and down from the tongue's back pellet position (T4). This anchoring might be improved in one of three ways: by eliminating the anchored points, by making the anchored points farther back and down, or by making the anchored points move when the tongue moves. Also, the tongue was not evaluated on whether or not it poked through the palatal outline. In all but a few cases, the tongue did poke through the palatal outline (particularly the alveolar ridge). Since this behavior did not affect the external appearance of face, it was not assigned a negative point.

Overall, in Table 5.3, the ratings for teeth position were 1 EXCELLENT, 9 VERY GOOD, 3 GOOD, and 2 FAIR, the ratings for lip position were 1 EXCELLENT, 4 VERY GOOD, 4 GOOD, 5 FAIR, and 1 VERY POOR and the ratings for tongue position were 2 EXCELLENT, 4 VERY GOOD, 6 GOOD, and 3 FAIR. The subject who appears to be animated best is JW502, with ratings of VERY GOOD, VERY GOOD, and EXCELLENT. The subject who appears to animated worst is JW11, with ratings of VERY POOR, FAIR,



JW19 JW32

Figure 5.6: Front View of JW19 and JW32 with Palatal Outline Extending below the Top Front Teeth



JW45 JW18 JW32 JW41

Figure 5.7: Side View of a Nicely Shaped Jaw (JW45) and Unusually Shaped Jaws (JW18, JW32, and JW41)

and VERY GOOD.

5.2.2 Overall Face Shape Evaluation

An evaluation, summarized by Table 5.4, was performed on the overall appearance of the initialized face. Negatives were assigned for an unusually placed palatal outline and for a strangely shaped or placed jaw. For instance, JW19 had a palatal outline that extended below the tip of the front upper teeth (see Figure 5.6). Because it was very noticeable, JW19 was assigned two negatives, giving this subject and rating of GOOD. A few subjects had unusually placed jaws (see Figure 5.7). For instance, JW18 had a very long jaw, JW32 had a very square jaw, and JW41 had an unusual shape to the back of his jaw. JW18 and JW41 were assigned one negative for their respective faults and were, thus, given a rating of VERY GOOD. JW32, who also had the palatal outline extending below the tip of the the front upper teeth, was assigned two negatives overall and a rating GOOD.

To summarize the subjective evaluation of the face shape, two thirds of the faces had EXCELLENT face shape, and the remaining one third had VERY GOOD (3) or GOOD (2) face shape. These ratings were made on the basis of jaw shape and palatal outline placement.

Subject	Overall Face Shape
JW11	EXCELLENT
JW12	EXCELLENT
JW15	EXCELLENT
JW16	EXCELLENT
JW18	VERY GOOD –long jaw
JW19	GOOD – –palatal outline extends below the tip of the front upper teeth
JW21	VERY GOOD –square jaw
JW24	EXCELLENT
JW25	EXCELLENT
JW27	EXCELLENT
JW32	GOOD –palatal outline extends below the tip of the front upper teeth –square jaw
JW40	EXCELLENT
JW41	VERY GOOD –jaw appears to be too large when viewed from the side
JW45	EXCELLENT
JW502	EXCELLENT

Table 5.4: Subjective Evaluation of Overall Face Shape

5.2.3 Discussion

The variety of ratings awarded to the subjects suggests that certain faces fit the original mesh better than others. For example, with regard to lip positioning, some subjects had lips that were too close together, and other subjects had lips that were too far apart. Possibly some subjects have thicker or thinner lips, but CASSI 1.0 makes all lips the same size and shape. Thinner lips should be used for those whose lips are held tightly together, and thicker lips should be used for those whose lips never touch.

Because the lips play a central role in speech, CASSI 2.0 and 2.1 were implemented to improve the shape and movement of the lips. In particular, varying lip thicknesses were created by "stretching or shrinking" the original lip mesh. The result is that each subject in CASSI 2.0 and 2.1 has a unique lip thickness.

5.3 A Subjective Evaluation of CASSI 2.0 and 2.1

The focus of CASSI 2.0 was on improving the lip movements from CASSI 1.0 by rounding the lips, preventing the lips from running into the surface of the teeth, and giving the lips thickness. The lips were rounded using an ellipse, the lips were prevented from running into the surface of the teeth by using a parabolic track, and the thickness of each lip was created by using half the average distance between the upper and lower lip pellets. CASSI 2.1 was implemented to improve on this lip thickness by assigning one third the average distance to the upper lip and two thirds the average distance to the lower lip for subjects who satisfied the criteria for adjustments, as described in Section 4.3.1.

The following two sections present a comparitive evaluation of CASSI 2.0 and 2.1. Section 5.3.1 presents an evaluation on the general lip placement and movement for the tasks in Table 5.1. Section 5.4 presents an evaluation of the lip movement for the two major tasks, tasks 118 and 106, used for tuning CASSI 2.0 and 2.1.

5.3.1 Lip Placement and Movement

This section presents an evaluation on the general lip placement and movement for the tasks in Table 5.1. The evaluation, summarized by Table 5.5, presents a comparison of lips for the three versions of CASSI. The first column lists the subject. The second column is copied from Table 5.3 and contains the results of the evaluation performed on the lips in CASSI 1.0. The third and fourth columns provide, respectively, the evaluation of the lips for CASSI 2.0 and CASSI 2.1.

The lips were evaluated in a manner similar to the lip evaluation for CASSI 1.0. The front view was examined at the starting and ending frame to determine if the lips lightly touched (desirable), were too far apart, or were too close together; the front view was also examined during the animation to determine if the lips were well centered in relation to the teeth and to determine if the lips passed through the surface of the teeth; lastly the side view was examined to see if the bottom lip protruded more than the top. Negatives were assigned for problems, and extra negatives were assigned for significant problems.

Since this evaluation was meant to compare the three versions of CASSI, if a marked improvement occurred from the previous to current versions, fewer negatives were assigned to the current version to reflect this improvement. An example is JW16, shown in Figure 5.8, who, in CASSI 1.0, had a large separation between the lips (denoted by two negatives). In CASSI 2.0, this separation is smaller than the previous version and is, thus, assigned one negative. Similarly, if the new version resulted in worse lip positioning and the rating indicated that the versions were the same, then a extra negative was assigned. JW19, shown in Figure 5.9, was an example of this; from CASSI 2.0 to 2.1, the lips went from being centered slightly low to being centered too high. Without adding negatives, both of these would have had a rank of VERY GOOD. However, in CASSI 2.1, JW19 was considered to look worse than in CASSI 2.0 and an additional negative was assigned to give CASSI 2.1 a rating of GOOD. Because of the implementation of CASSI 2.1, sometimes there was no change in the lip thickness between 2.0 and 2.1. The subjects whose lips remain at half the average distance are denoted in column four by "no change from CASSI 2.0" and the rating remains the same.

Figures 5.8 and 5.9 contain snapshots taken from the initial frame of animation and are meant to demonstrate some of the changes, noted in Table 5.5, that occur in lip thickness. Figure 5.8 demonstrates the changes from CASSI 1.0 to CASSI 2.0, and Figure 5.9 demonstrates the changes from CASSI 2.0 to CASSI 2.1. Both of these figures are arranged in a tabular format. The first column lists the subject and task number separated by an underscore. The second and third columns provide snapshots of the initial frame of the animation for each of the two versions being compared.

Figure 5.8 demonstrates a few of the differences between CASSI 1.0 and 2.0. The snapshots in the first row demonstrate the improvement in JW11's lip thickness; the lips are changed from having a large separation to being lightly placed together. The second row demonstrates the improvement in JW16. Although, a slight fault, a gap between the lips, still exists in CASSI 2.0, this is a marked improvement over CASSI 1.0. The final row shows JW15, which undergoes no change in lip thickness from CASSI 1.0 to CASSI 2.0; the slight gap between the lips does not change.

Figure 5.9 shows the differences between CASSI 2.0 and 2.1. The first row shows JW19 with lips that are centered slightly low in CASSI 2.0 and lips that are centered too high in CASSI 2.1. The right-hand snapshot in the first row demonstrates that the lower teeth are rarely seen in CASSI 2.1. The second and third rows show the improvement in the lip centering for JW502 and JW12, respectively. The snapshots demonstrate that the lips are centered too low in CASSI 2.0 and are centered well in CASSI 2.1; similar portions of teeth are displayed within the inner contours of the two lips.

Table 5.6 summarizes the ratings of the lip movement and placement given in Table 5.5. The first column gives the rating. The second, third, and fourth columns give a count of the ratings assigned to respectively CASSI 1.0, 2.0, and 2.1. For instance, CASSI 2.0 had 0 VERY POOR, 1 POOR, 5 FAIR, 3 GOOD, 6 VERY GOOD, and 0 EXCELLENT. Although CASSI 2.0 provided lip rounding, which was not analyzed in the evaluation in Table 5.5, there was little subjective improvement from CASSI 1.0 to CASSI 2.0. There was, however, large improvement from CASSI 2.0 to CASSI 2.1; the ratings changed from 9 POOR/FAIR/GOOD to 5 POOR/FAIR/GOOD, and 6 VERY GOOD/EXCELLENT to 10 VERY GOOD/EXCELLENT.

Subject	CASSI 1.0	CASSI 2.0	CASSI 2.1
JW11	VERY POOR –lips go through surface of teeth – –there is a large separation between lips (they never touch) – –lips are centered low (often go below the bottom of the lower front teeth)	GOOD +lips lightly touch at beginning –lips appear too large when moving (overlap) –lips are centered low (never see top teeth)	VERY GOOD +lips are well centered –lower lip is too large (rarely see bottom teeth)
JW12	GOOD – –lips are centered low (go below the bottom of the lower front teeth)	FAIR +lips lightly touch at beginning – – –lips are centered low (go below the bottom of the lower front teeth)	VERY GOOD –lips are centered better than in other versions, but still go below the bottom of the lower front teeth
JW15	GOOD –lips are separated at beginning –lips go below the bottom of the lower front teeth	POOR –lips are separated at beginning – –lips are centered low (go below the bottom of the lower front teeth) –lips overlap in task 10	GOOD –lips are centered better than in CASSI 2.0, but still go below the bottom of the lower front teeth +overlapping lips are handled –lips are separated at beginning
JW16	FAIR – –there is a large separation between lips –lips are centered high	VERY GOOD -lips are separated at beginning	EXCELLENT no change from CASSI 2.0
JW18	GOOD –lips are tightly placed together –lips are centered high	VERY GOOD +lips lightly touch at beginning –lips are centered high	VERY GOOD no change from CASSI 2.0
JW19	GOOD – –there is a large separation between lips	VERY GOOD +lips lightly touch at beginning –lips are centered slightly low	GOOD –lips are centered high (rarely see bottom teeth) –looks worse than CASSI 2.0
JW21	VERY GOOD –lips go through surface of teeth	VERY GOOD +lips lightly touch at beginning (slight gap) –lips are centered low (go below the bottom of the lower front teeth)	EXCELLENT +lips are well centered
JW24	FAIR +from front view, lips are nicely closed (for task 18) – –from side view, lower lip protrudes farther than upper lip –lips are well centered but go below the bottom of the lower front teeth	FAIR +lips lightly touch at beginning (for task 18) – –from side view, lower lip protrudes out farther then upper lip –lips are centered low (go below the bottom of the lower front teeth)	FAIR no change from CASSI 2.0
JW25	FAIR –lips are tightly placed together –upper lip goes through the surface of the upper set of teeth –from side view, lower lip protrudes out farther than upper lip	FAIR –from side view, lower lip protrudes out farther than upper lip –lips are separated at beginning –lips are centered low (go below the bottom of the lower front teeth)	FAIR no change from CASSI 2.0
JW27	VERY GOOD –lips are tightly placed together	FAIR –lips are separated at the beginning – –lips are centered low (go below the bottom of the lower front teeth)	GOOD –lips are separated at the beginning –lips are well centered, but still go below the bottom of the lower front teeth +lips are well centered (although they still go below the lower front teeth)
JW32	FAIR –lips go through the surface of teeth –from side view, lower lip protrudes out farther than upper lip (slight) –lips are tightly placed together at beginning	GOOD +lips lightly touch at beginning –from side view, lower lip protrudes farther than upper lip (slight) –lips are centered low (go below the lower front teeth)	VERY GOOD +lips are well centered –from side view, lower lip protrudes out farther than upper lip (slight)
JW40	VERY GOOD –there is a separation between lips at beginning	VERY GOOD +lips lightly touch at beginning –lips are centered low (never see top teeth)	EXCELLENT +lips are well centered
JW41	FAIR –there is a separation between lips at beginning – –lips are centered low (go below the bottom of the lower front teeth)	FAIR +lips lightly touch at beginning (of task 11) – –lips are centered low (go below the bottom of the lower front teeth) –lips overlap	EXCELLENT +lips are well centered +overlapping lips are handled
JW45	EXCELLENT	VERY GOOD +lips lightly touch at beginning –lips are centered low (never see top teeth)	EXCELLENT +lips are well centered
JW502	VERY GOOD –lips are tightly placed together	GOOD –there is a slight separation between the lips –lips are centered low (go below the bottom of the lower front teeth)	VERY GOOD +lips are well centered –there is a slight separation between the lips

Table 5.5: Subjective Evaluation of Lip Movement and Initial Placements in CASSI 1.0, 2.0 and 2.1

Rating	CASSI 1.0	CASSI 2.0	CASSI 2.1
VERY POOR	1	0	0
POOR	0	1	0
FAIR	5	5	2
GOOD	4	3	3
VERY GOOD	4	6	5
EXCELLENT	1	0	5

Table 5.6: Summary of Lip Ratings for CASSI 1.0, 2.0, and 2.1

Subj_Task	CASSI 1.0	CASSI 2.0
JW11_019		
JW16_006		
JW15_009		

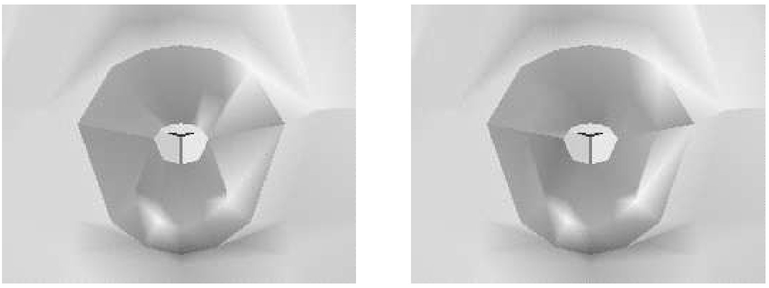
Figure 5.8: Differences between CASSI 1.0 and CASSI 2.0

Subj_task	CASSI 2.0	CASSI 2.1
JW19_001		
JW502_021		
JW12_002		

Figure 5.9: Differences between CASSI 2.0 and CASSI 2.1

Subject	Task 118?	Task 106?
JW11	no	yes (0)
JW12	no	no
JW15	no	no
JW16	yes (0)	no
JW18	yes (6)	no
JW19	yes (5)	no
JW21	no	no
JW24	no	yes (4)
JW25	yes (4)	no
JW27	yes (0)	yes (0)
JW32	no	no
JW40	yes (3)	yes (2)
JW41	yes (2)	no
JW45	yes (0)	yes (3)
JW502	yes (1)	yes (1)

Table 5.7: Subjects with Tasks 118 and 106



(a)Before **(b)After**

Figure 5.10: Rounded Lips for JW16, before and after Adding Vertex 300

5.4 Maximum Lip Protrusion and Jaw Wagging

To tune CASSI 2.0 and 2.1, both the jaw wagging and maximum lip protrustion task were animated. The maximum lip protrusion task was examined to ensure that the lips rounded smoothly forward without any large jumps or changes from frame to frame, especially at the threshold where the lips start rounding. The jaw wagging task was examined for smooth motion, with particular attention paid to when the lips change from a horizontal to a vertical ellipse. Once satisfactory motion had been created in these two tasks, the other tasks were examined to determine if further adjustments were required.

As mentioned in the previous paragraph, the two major tasks used for tuning were the maximum lip protrusion task, task 118, and the jaw wagging task, task 106. For some subjects, these task files were not usable due to mistracking errors or missing pellet coordinate values. Table 5.7 indicates whether or not a particular subject has usable files for tasks 118 and 106. The first column lists the subject. The second and third column each contain a "yes" or a "no" based on whether the subject had usable files for task 118 and task 106, respectively. For all "yes" answers in the column, the number written in parenthesis identifies the order that the subject's task file was tested. At the start, only a small set of tasks was tested. After the performance on those tasks improved, other tasks were added to the testing. For instance, with task 118, testing was started on three subjects: one with subjectively large-sized lips (JW16), one with subjectively small-sized lips (JW27), and one with subjectively medium-sized lips (JW45). These three subjects are denoted in Table 5.7 with a zero in parenthesis. Once satisfactory movement of the lips was obtained, the other subjects were incrementally added in the following order: JW502, JW41, JW40, JW25, JW19, and JW18. Task 106 was tested with this similar incremental approach.

At first, the lips were rounded using the original topology and vertices inherited from Parke's code.

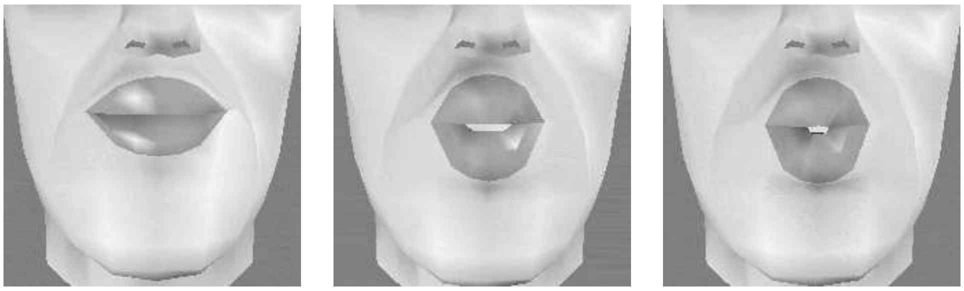


Figure 5.11: Front View of Lip Protrusion Task for JW16 in CASSI 2.0

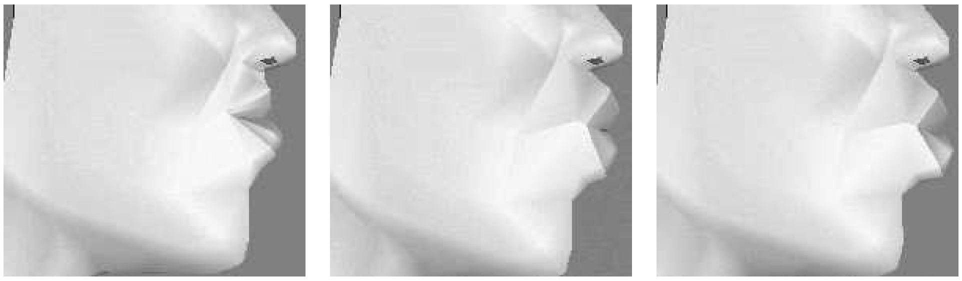


Figure 5.12: Side View of Lip Protrusion Task for JW16 in CASSI 2.0

Unusual shapes and shadows, as shown in Figure 5.10(a), resulted. To create lips with a smoother surface, an additional corner point, vertex 300, was added between the two existing corner vertices, 70 and 287. Figure 5.10(b) shows the smoother shape of the rounded lips that resulted.

Figure 5.11 and 5.12 show three frames of the animation created by JW16's maximum lip protrusion task (task 118). These figures demonstrate the movement of the lips during lip protrusion, as seen from the front and side views. Figure 5.11 shows the front view of the lips as they move from comparitively wide ellipses to rounded ellipses, based on the increasing protrusion of the lips. Figure 5.12 shows the side view of the lips demonstrating the forward movement of the lips and the bulging in the lips that occurs when they are brought forward.

Figure 5.13 and Figure 5.14 provide front and side views, respectively, of task 106 (the jaw wagging task) as performed by subject JW27 in CASSI 2.0. These figures are meant to demonstrate the movement of the lips and jaw, which have coordinated up and down motion even though their parameters are adjusted separately.

Because CASSI 2.0 and 2.1 were highly influenced by tasks 118 and 106, two separate evaluations were made of these two tasks. These evaluations are summarized by two tables, Table 5.8 and Table 5.9. The fol-

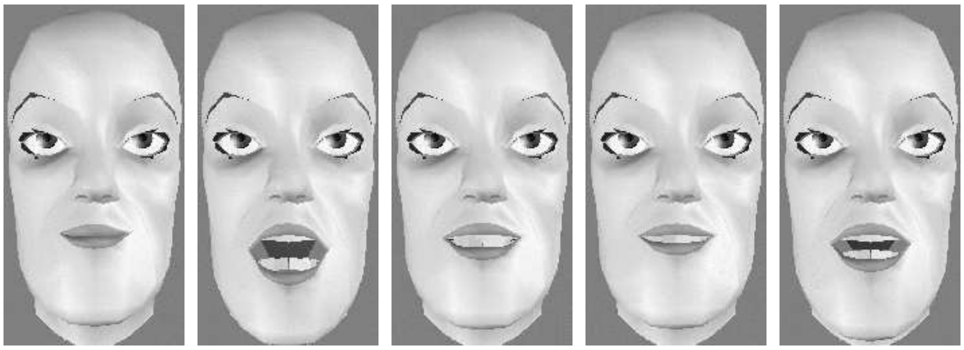


Figure 5.13: Front View of Jaw Wagging Task for JW27 in CASSI 2.0

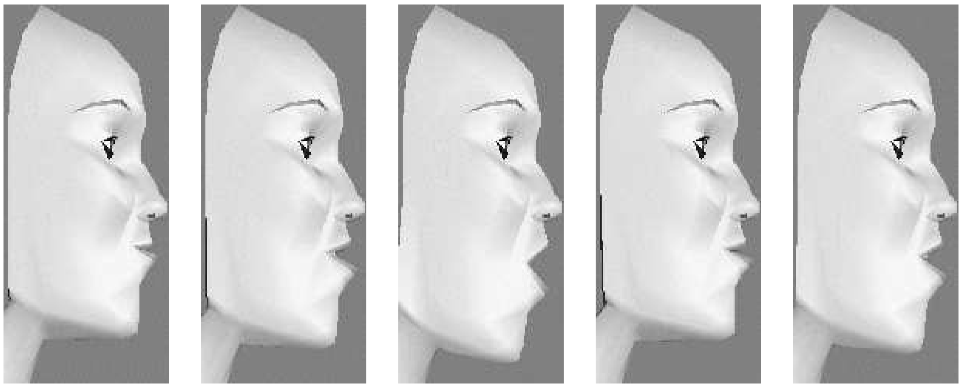


Figure 5.14: Side View of Jaw Wagging Task for JW27 in CASSI 2.0

lowing two subsections present further details on these tables, summarizing the evaluations on the maximum lip protusion and jaw wagging tasks.

Lip Rounding

To evaluate the lip rounding movement, the maximum protrusion task, task 118, was used. The results of this evaluation are summarized in Table 5.8. The first column lists the subject; only subjects with task 118, as identified in Table 5.7, are listed in this first column. The second, third, and fourth columns summarize the evaluation of lip protrusion from the front and side views for CASSI 1.0, CASSI 2.0, and CASSI 2.1, respectively. The rating system was the same one that was applied to Tables 5.3 and 5.5.

Results for CASSI 1.0 are given in the first column of Table 5.8. As shown by Figure 5.15, this version did not create rounded lips. The snapshots shown in Figure 5.15 (a) and (c) were created by CASSI 1.0, and the snapshots shown in Figure 5.15 (b) and (d) were created by CASSI 2.0. These snapshots demonstrate that the lips move up and down and that the corners are never drawn together or forward. Because CASSI 1.0 did not have lip rounding, each subject in this column was, by default, assigned three negatives for "no lip rounding". Negatives were also assigned in this column for the following reasons: lips that ran into the nose at maximum protrusion, points above the the lips that "caved" in when the lips were brought upwards, lips that were centered too high or too low, and lips that stick out too much from the chin or from each other at maximum protrusion.

For the other two versions, CASSI 2.0 and 2.1, the front view was examined at maximum protrusion to determine if the rounded lips had any anomolies. JW41 and JW19 in Figure 5.16 demonstrate two undesirable anomolies in the front view. JW41 shows lips that are too big, and JW19 shows lips that run into the surface of the nose. JW16 is included in Figure 5.16 to represent lips that are nicely shaped. The front view was also examined during the animation to determine if the lips were centered too low or too high (undesirable).

The side view was examined to determine if, at the initial frame, the lower lip protruded considerably more than the upper lip; and also to determine if, at maximum protrusion, the lips protruded too much from the chin, creating unnatural shaped lips. Figure 5.17 demonstrates this unnatural lip protrusion. The first snapshot shows JW16, whose lips appear well placed at maximum protrusion, and the second snapshot shows JW18, whose lips stick out unnaturally from the chin.

From both the front and side view, the lips were also examined for "jitter", which occurred when the lips appeared to shake in and out slightly. All subjects had slight jitters of the lip. However, the jaw also had the same amount of jitter. Because the jitter was consistent with the X-ray microbeam data, it was not considered to be a fault in CASSI.

Table 5.8 also compares CASSI 2.0 to 2.1. In the table, some subjects were marked with the comment, "no change", and the rating remained the same. Because of the implementation, the distribution in the lip thickness for that particular subject did not change from CASSI 2.0 to 2.1. With CASSI 2.1, the centering for most lips is improved, but a different problem occurs; from the side view, the upper lip has vertices which appear to be unusually placed, creating shadows and an abnormal shape to the profile outline of the

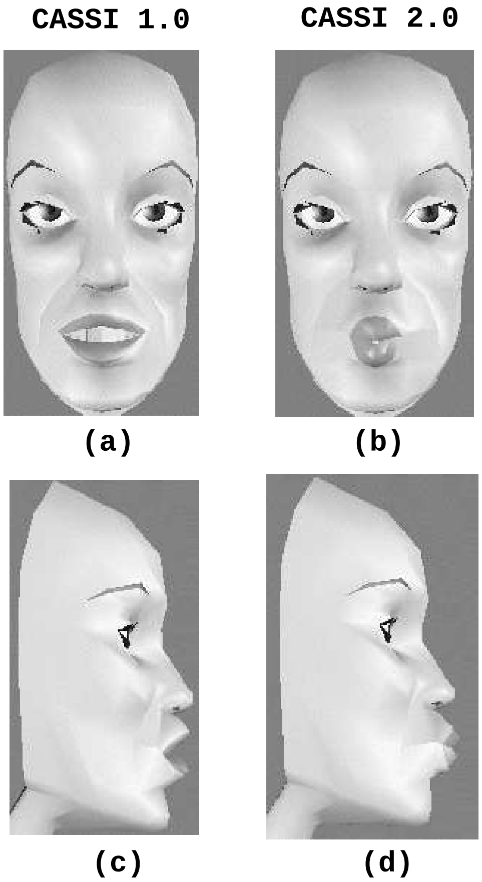


Figure 5.15: Maximum Protrusion Task with CASSI 1.0 and 2.0

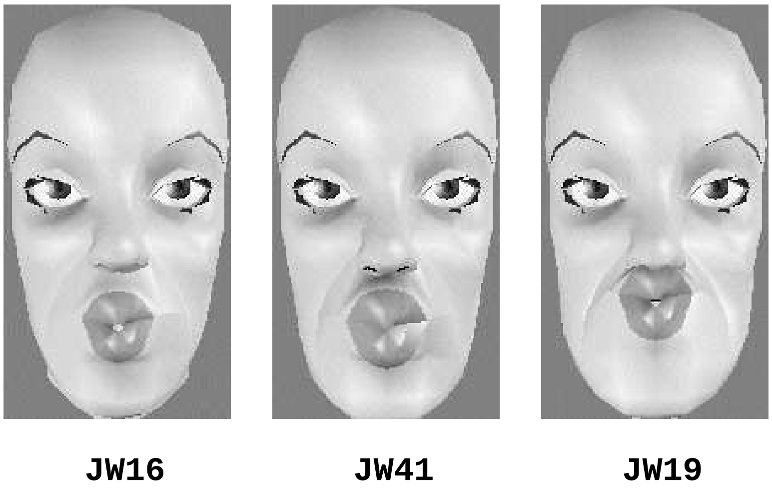
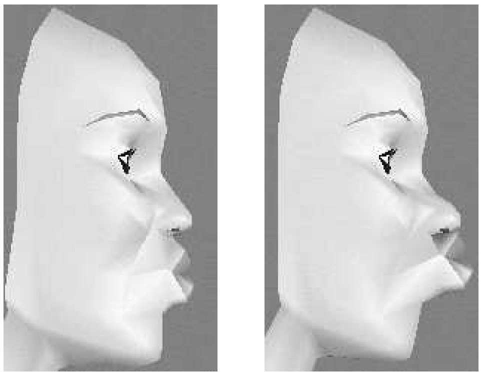


Figure 5.16: Front View of Good and Faulty Lip Rounding in CASSI 2.0

Subject	CASSI 1.0	CASSI 2.0	CASSI 2.1
JW16	FAIR – – –no lip rounding	EXCELLENT +good lip shape	EXCELLENT no change
JW18	UNACCEPTABLE – – –no lip rounding –from side view, lips stick out too much from chin at max protrusion –from side view, points above the lip cave in –lips are centered high	FAIR –from side view, lips stick out too much from chin at max protrusion –lips run into nose at max protrusion –lips too big at max protrusion	FAIR no change
JW19	UNACCEPTABLE – – –no lip rounding – –lips run into nose at max protrusion – –from side view, points above the lip cave in –lips are centered high	GOOD – –lips run into nose at max protrusion	FAIR – –lips run into nose at max protrusion –lips are centered too high
JW25	POOR – – –no lip rounding –from side view, lower lip protrudes in initial position	GOOD –from side view, lower lip protrudes in initial position –lips are centered low	GOOD +lips are better centered –from side view, lower lip protrudes in initial position –from side view, upper lip has sharp profile
JW27	VERY POOR – – –no lip rounding –from side view, lips stick out too much from chin at max protrusion –lips are centered low	GOOD –from side view, lips stick out too much from chin at max protrusion –lips are centered low	GOOD +lips are better centered –from side view, lips stick out too much from chin at max protrusion –from side view, upper lip has sharp profile
JW40	VERY POOR – – –no lip rounding –from side view, top lip protrudes more than bottom at max protrusion –lips are centered low	GOOD –from side view, top lip protrudes more than bottom at max protrusion –lips are centered low	EXCELLENT +lips are better centered +side view improved (top lip does not stick out as much)
JW41	VERY POOR – – –no lip rounding –from side view, lips stick out too much from chin at max protrusion (chin is odd shape) –lips are centered low	GOOD –lips too big at max protrusion –lips are centered low	GOOD +lips are better centered –lips too big at max protrusion –from side view, upper lip sometimes has sharp profile
JW45	POOR – – –no lip rounding –lips are centered low	VERY GOOD –lips are centered low	VERY GOOD +lips are better centered –from side view, upper lip has sharp profile
JW502	FAIR – – –no lip rounding	VERY GOOD –lips are centered low	VERY GOOD +lips are better centered –from side view, upper lip has sharp profile

Table 5.8: Subjective Evaluation of Maximum Lip Protrusion Task



JW16 **JW18**

Figure 5.17: Side View of Good and Faulty Lip Rounding in CASSI 2.0

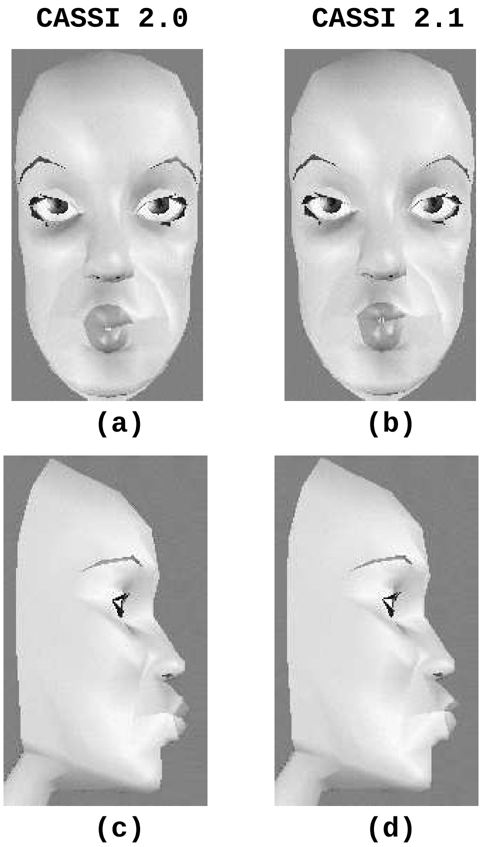


Figure 5.18: Differences in Lip Rounding for Subject JW45

lips. Figure 5.18 demonstrates the change in lip thickness from CASSI 2.0 to CASSI 2.1. The front views, Figure 5.18(a) and (b), show that the upper lip is thinner with CASSI 2.1, which creates better centered lips. The side views, Figure 5.18(c) and (d), also show the thinner upper lip in CASSI 2.1 and show the profile of the lips with sharper edges.

One inconsistency can be noted for JW25 in Table 5.5 and Table 5.8. In Table 5.5, JW25 was evaluated based on tasks 15 and 16, and the thickness of the lips did not change from CASSI 2.0 to 2.1. In Table 5.8, JW25 was evaluated based on task 118, and the thickness of the lips did change from CASSI 2.0 to 2.1. This inconsistency, from the unchanged to changed thickness of the lips, can be explained by the different task files used. Along with the criteria for modifying the lip thickness is an underlying assumption that the lips will start in the same position at the beginning of each task file. For JW25, however, the lips have a different starting location for task 118 than for tasks 15 and 16. Specifically, at the starting location for task 118, the lips exceed the threshold that determines if the lips are too far over the edge of the top front teeth, and the thickness of the lips is adjusted. For the other two tasks, the lips are not positioned sufficiently over the edge of the top teeth to exceed the threshold, and thus, for these two cases, the thickness of the lips is not adjusted.

Table 5.8 reflects the marked improvement from CASSI 1.0 to CASSI 2.0 due to the addition of lip rounding. Table 5.8 also shows that there is little change from CASSI 2.0 to 2.1. The ratings for CASSI 1.0 are distributed as follows: 2 FAIR, 2 POOR, 3 VERY POOR, and 2 UNACCEPTABLE. For CASSI 2.0, the ratings are distributed as follows: 1 EXCELLENT, 2 VERY GOOD, 5 GOOD, and 1 FAIR. For CASSI 2.1, the ratings are: 2 EXCELLENT, 2 VERY GOOD, 3 GOOD, and 2 FAIR. In summary, the ratings are as follows: 2 FAIR and 7 POOR/VERY POOR/UNACCEPTABLE, in CASSI 1.0; 3 EXCELLENT/VERY GOOD and 6 GOOD/FAIR in CASSI 2.0; and 4 EXCELLENT/VERY GOOD and 5 GOOD/FAIR, in CASSI 2.1. Although CASSI 2.0 and 2.1 provide very similar quality animation for lip rounding, these versions provide an improvement from CASSI 1.0, which does not provide lip rounding.

Jaw Wagging

The jaw wagging task, task 106, was also used to tune CASSI 2.0 and 2.1. While performing task 106, the movements of the jaw, chin, and lips were viewed and evaluated from both the front and side views. The results are summarized in Table 5.9. The first column lists all the subjects with usable task 106 files. The second, third and fourth column evaluate the jaw wagging movement in CASSI 1.0, 2.0, and 2.1, respectively.

Negatives were assigned to the front view in the following cases: (1) if the lips were very triangular, or diamond-shaped, particularly, when the lips were opened wide, and (2) if the lower lip passed through the bottom set of teeth. Negatives were assigned to the side view in the following cases: (1) if the points under the lip "caved-in" and created unusual shapes and shadows, (2) if the corners of the lip were not well positioned in relation to the other points of the lip, (3) if the chin or lower jaw "caved-in" and created unusual shapes and shadows, and (4) if the corner of the lips jittered, or there was a large frame jump, in the transition from horizontal to vertical ellipse. Extra negatives were assigned to problems that were especially predominate.

Figure 5.19 shows some of these problems in CASSI 1.0. The figure is in tabular format. The first column lists the subject and the view. The second and third column show two corresponding frames for the jaw wagging task created by CASSI 1.0 and 2.0, respectively. The first two rows show the front view of subjects JW11 and JW45, respectively. The snapshots on the left show problems associated with JW11 and JW45 in CASSI 1.0: the lower lip goes through the surface of the teeth, and the lips have a diamond shape; and the snapshots on the right show that these two problems have been alieviated in CASSI 2.0. The last two rows show the side view of subjects JW45 and JW502, respectively. The problems with these two subjects in CASSI 1.0 are as follows: the points under the lip cave in, the corner point is not well placed, and the chin caves in making a strange shape, especially seen in the snapshot of JW45. Again, as demonstrated by the snapshots on the right, these problems have been alieviated in CASSI 2.0.

Subjectively, the shape of the lips is improved from CASSI 1.0 to 2.0, and from CASSI 2.0 to 2.1 for the jaw wagging task. The following can be said: from the front view, CASSI 2.0 no longer has a triangular-shaped upper lip and the teeth no longer go through the surface of the teeth; and from the side view, CASSI 2.0 has improved the motion of the points under the lip, the motion of the corner points, and, for the most part, the motion of the jaw, or chin. CASSI 2.1 has improved the motion of the lips so that there is no

Subject	CASSI 1.0	CASSI 2.0	CASSI 2.1
JW11	UNACCEPTABLE –lips are diamond shaped when wide open(front-view) – –lower lip goes through surface of teeth – –points under lip cave in (side view) – –corner point is not well placed	GOOD – –from the side view, the corner of the lips jitter in transition from horizontal to vertical ellipse	EXCELLENT
JW24	GOOD –lips are diamond shaped when wide open(front-view) –points under lip cave in (side view)	VERY GOOD –from the side view, the corner of the lips jitter in transition from horizontal to vertical ellipse	EXCELLENT
JW27	POOR –lips are diamond shaped when wide open(front-view) –points under lip cave in (side view) –corner point is not well placed –chin caves in making strange shapes	VERY GOOD –from the side view, the corner of the lips jitter in transition from horizontal to vertical ellipse	EXCELLENT
JW40	UNACCEPTABLE –lips are diamond shaped when wide open(front-view) – –points under lip cave in (side view) – –corner point is not well placed –chin caves in making strange shapes	FAIR – –from the side view, the corner of the lips jitter in transition from horizontal to vertical ellipse –chin caves in making strange shapes	VERY GOOD –chin caves in making strange shapes
JW45	UNACCEPTABLE –lips are diamond shaped when wide open(front-view) –lower lip goes through surface of teeth – –points under lip cave in (side view) – –corner point is not well placed –chin caves in making strange shapes	GOOD – –from the side view, the corner of the lips jitter in transition from horizontal to vertical ellipse	EXCELLENT
JW502	UNACCEPTABLE – –lips are diamond shaped when wide open(front-view) – –lower lip goes through surface of teeth – – –points under lip cave in (side view) – – –corner point is not well placed – –chin caves in making strange shapes	FAIR – –from the side view, the corner of the lips jitter in transition from horizontal to vertical ellipse –chin caves in making strange shapes	VERY GOOD –chin caves in making strange shapes

Table 5.9: Subjective Evaluation of Jaw Wagging Task

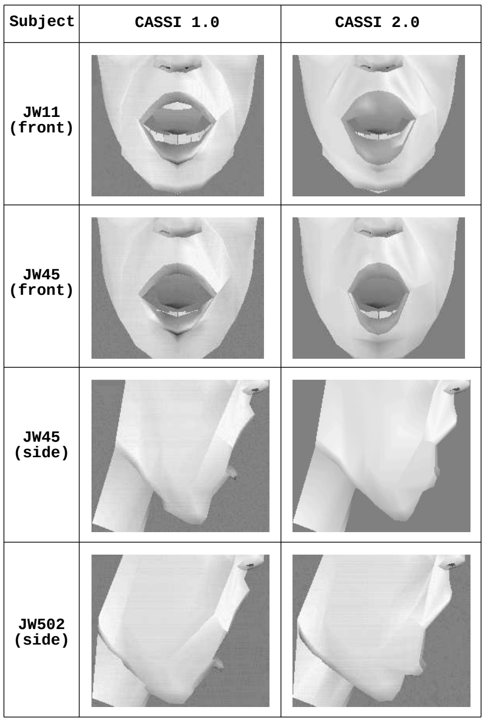


Figure 5.19: Differences in Jaw Wagging Task from CASSI 1.0 to 2.0

longer a jitter in the transition from horizontal to vertical ellipse. These improvements are reflected in the ratings for each of these versions. CASSI 1.0 had the following ratings: 4 UNACCEPTABLE, 1 POOR, and 1 GOOD, CASSI 2.0: 2 FAIR, 2 GOOD, and 2 VERY GOOD, and CASSI 2.1: 2 VERY GOOD, and 4 EXCELLENT.

Chapter 6

Conclusions and Future Research

6.1 Summary

Several techniques are used to create facial animations, including image-based key frame animation, parametric key frame animation, performance based animation, and speech synchronized animation. This research focuses on one approach, which can be best categorized as a performance based animation that augments Parke's model.

Our system, called CASSI (Computer Animated Speech SImulator), makes use of Parke's model, a parametric key frame model. As a parametric key frame model, the movement of the face of Parke's model is controlled by specifying a few parameter values, rather than by specifying the entire image frame. This model was chosen for the following reasons: because of its simplicity and low computational complexity (when compared, specifically, to the muscle-based model), because the source code was easily accessible, and because several researchers interested in the correspondence between the auditory signal and the animation are using Parke's model [9] [6] [16].

Although Parke's model is a parametric key frame model, our approach cannot be classified as a parametric key frame animation technique. By definition, parametric key frame animation involves two parts: (1) determining and setting parameter values for a desired key frame, and (2) interpolating parameter values to smooth between these key frames. Our approach does not involve the typically manual work of determining and setting parameter values, and it does not require interpolation. Rather, the parameter values are automatically set for each frame of animation, and interpolation is not needed since each frame, in essence, becomes a "key" frame.

Our approach is, however, categorized as a performance-based animation technique because human motion is used to drive our animations. The human motion is encoded as 2D X-ray microbeam data, which track the side view movement of gold pellets placed on the lips, tongue, and jaw of several subjects. These pellets were tracked for each subject while he or she performed different tasks. The coordinate system, developed to have a common anatomically based reference frame for all speakers, has the origin placed between the tips of the top front teeth.

The major achievement described in this document was developing an interface that moved the lips, tongue, teeth and jaw of the 3D facial model according to the movements of the 2D X-ray microbeam pellets. For our approach, several changes were made to Parke's original set of parameters; some parameters were modified or disabled, and some parameters were added. For example, the functionality of the jaw rotation parameter and its associated procedure were adjusted for integration with the X-ray microbeam data. New parameters for lip and tongue movement were added. Parameters associated with the lips, teeth, and chin were disabled so that they did not interfere with the new parameters and initializations. Lip rounding was also added to Parke's model, creating more diverse lip movement.

Our implementation involved three versions. CASSI 1.0 focused on the initialization and animation of the lips, teeth, jaw, and tongue. CASSI 2.0 and 2.1 were implemented to improve the lip motion. Each version provided major contributions. Once completed, the version underwent a subjective evaluation to determine where improvements were needed as well as where improvements had been made.

CASSI 1.0 provided initialization and animation of the lips, teeth, jaw, and tongue. Its contributions included creating a correspondence between the pellets of the XRMB data and specific vertices in Parke's model, initializing the positions of the tongue, lips, teeth, and jaw of the 3D face, and animating these components of speech using new parameters and a modified version of Parke's jaw rotation procedure. The most significant contribution of this version was the modified jaw rotation procedure, which uses a point of rotation calculated frame by frame rather than using a fixed point.

CASSI 2.0 and 2.1 improved the movement and shape of the lips. The major contributions of CASSI 2.0 included creating lip rounding using elliptical outlines, preventing the lips from intersecting with the teeth using a parabolic track, and creating lip thicknesses based on each subject. To create lip rounding and lip thickness, data were extracted from the XRMB data. These data served two purposes: for lip rounding, they provided a reference for rounded lips based on resting and maximum forward and backward positions; and for lip thickness, they provided an estimate of the combined thickness of both lips. In CASSI 2.0, the thickness of one lip was half the combined thickness of both lips. In CASSI 2.1, subjects whose lips were centered too low, according to a threshold, were modified so that the upper lip was one third and lower lip was two thirds of the combined thickness of both lips. With these adjustments to lip thickness, CASSI 2.1 required blending functions to create a smooth transition from stretched to rounded lips for the corner points.

Two evaluations were performed on CASSI 1.0. The first evaluation focused on the initialization and movement of the lips, teeth, and tongue. The second evaluation focused on the overall face shape. The results of the first evaluation performed on the lips, teeth, and tongue suggest that EXCELLENT or VERY GOOD animations were produced about half the time, and VERY POOR, FAIR, or GOOD animations were produced about half the time. The results of the second evaluation suggest that the overall face shape was EXCELLENT two thirds of the time and VERY GOOD or GOOD one third of the time.

Three other evaluations were performed across the three versions of CASSI, CASSI 1.0, 2.0, and 2.1. The first evaluation focused on the lip placement and movement produced by the tasks given in Table 5.1. The second evaluation focused on the lip rounding produced by the maximum lip protrusion task, task 118. The third evaluation focused on the jaw wagging motion created by task 106. The following summarizes the results. Based on the evaluation of lip placement and movement, little improvement was made from CASSI 1.0 to CASSI 2.0; however, large improvement was made from CASSI 2.0 to CASSI 2.1. Based on the lip rounding evaluation, there was a marked improvement from CASSI 1.0 to CASSI 2.0 because CASSI 2.0 included lip rounding, while CASSI 1.0 did not. This same evaluation suggested that little further improvement occurred from CASSI 2.0 to 2.1. Based on the jaw wagging evaluation, the lips improved both from CASSI 1.0 to CASSI 2.0 and from CASSI 2.0 to CASSI 2.1. In summary, the results suggest that, the motion and placement of the lips improved from CASSI 1.0 to 2.1.

6.2 Conclusions

This research demonstrated that facial animation can be driven by X-Ray Microbeam data. The approach devised for this research and implemented in the CASSI software system is best classified as a performance-based approach using an extended Parke's parametric model. As mentioned in Chapter 1, this research also provides three original contributions: (1) it models and animates a simple tongue, which other performance based approaches do not provide because the recording techniques used, such as video taping, cannot capture the movement of the tongue; (2) several subjects are handled without requiring manual adjustments; and (3) it translates 2D sideview movement into 3D facial movement, in particular, it handles the frontview movement of the lips. The specific subjective evaluation technique used to evaluate the facial animations is also original to this research and, as such, provides a minor contribution.

According to our subjective evaluation, the quality of the animations produced by the software improved from CASSI 1.0 to 2.1, specifically, with regard to the lip movement. CASSI 1.0 provides excellent or very good animations about half the time, and very poor, fair, or good animations about half the time. The evaluations of CASSI 2.0 and 2.1 demonstrate that the lip movement and placement improved from CASSI 1.0 to CASSI 2.1, which, in turn, increased the number of excellent or very good animations.

6.3 Future Research

Additional research could be done on related topics. To improve the overall appearance of the face, the model could be augmented with hair, facial expressions, such as nodding and eye blinking, and texture mapping. Since an important part of this work is the inclusion of a tongue, this component could be improved by giving it thickness and curve; this change would improve the current, flat surface implementation. Another suggestion for future work would be to further adjust the 3D facial model for each subject's face. For instance, the width of the face could be adjusted according to the data provided in *Headmeasures.txt* for each subject. On a much larger scale, speech synchronization could be implemented. Speech synchronization would provide additional feedback on the realism of the animations created. As well, combining the sound file to the pellet coordinates file would serve as a stepping stone for the correspondence table approach to speech-driven animation, which was described in Section 1.1.

In addition, other possible future work includes fixing the problems noted in the subjective evaluation, specifically, in Section 5.2. Possible changes include: (1) adding collision detection to ensure that the tongue does not poke through the teeth or the palatal outline; (2) adjusting the anchored points in the tongue so these points move along with the rest of the tongue, are eliminated, or are moved farther back and down; and (3) adjusting the palatal outline so that it does not extend below the tip of the top front teeth in any subject.

Acknowlegements

We would like to thank Carl Johnson and the other members of the X-ray microbeam group at the University of Wisconsin for providing us with data, as supported by research grant number R01 DC 00820 from the National Institute on Deafness and Other Communicative Disorders, U.S. National Institutes of Health (X-ray Microbeam data). We thank Bruce Gilligan, Chris Shaw, and Xue-Dong Yang for comments. We acknowledge the Natural Sciences and Engineering Research Council of Canada, which supported this research by means of a Postgraduated Scholarship (Scheidt) and a Research Grant (Hamilton).

Nova would like to thank the many people who contributed to this research. This includes: Jen Turnbull for the CASSI name; Ben Korvemaker, Terry Peckham, Dee Jay Randall, and Chris Shaw for their ideas and comments; Kevin Munhall (Queen's University) for personally demonstrating his research; and Peter Alfonso (Indiana University), Jonas Beskow (Centre for Speech Technology, Sweden), Martin Cooke (University of Sheffield), Sorin Dusan (University of Waterloo), Vincent Gracco (Haskins Labs), Athanassios Hatzis (University of Sheffield), Masaaki Honda (NTT Basic Labs, Japan), John Mason (University of Wales Swansea), and Alan Wrench (Queen Margaret College, Edinburgh), for replying to the emails regarding background research on facial animation or articulatory data.

Bibliography

[1] P. Allard, I. A. F. Stokes, and J.-P. Blanchi. *Three-Dimensional Analysis of Human Movement.* Human Kinetics, Windsor, ON, 1995.

[2] H. Anton. *Calculus with Analytic Geometry, Third Edition.* John Wiley & Sons, Toronto, 1988.

[3] Synopsis of facial animation in *Antz*, November 1998, from http://www.antz.com/technology/technology_face2.html.

[4] P. Bergeron and P. Lachapelle. Controlling facial expressions and body movements in the computer-generated animated short "tony de peltrie". In *Advanced Computer Animation, SIGGRAPH '85 Tutorials*, volume 10, pages 61–79. ACM, New York, 1985. Synopsis and Picture obtained from http://mambo.ucsc.edu/psl/bergeron.html.

[5] J. Beskow. Rule-based visual speech synthesis. In *Proceedings of Eurospeech '95*, Madrid, Spain, September 1995. Paper obtained from http://www.speech.kth.se/multimodal/papers/.

[6] J. Beskow. Animation of talking agents. In *Proceedings of AVSP '97, ESCA Workshop on Audio-Visual Speech Processing*, Rhodes, Greece, September 1997. Paper obtained from http://www.speech.kth.se/multimodal/papers/.

[7] J. Beskow, M. Dahlquist, B. Granstrom, M. Lundeberg, K.-E. Spens, and T. Ohman. The teleface project multi-modal speech-communication for the hearing impaired. In *Proceedings of Eurospeech '97*, Rhodes, Greece, September 1997. Paper obtained from http://www.speech.kth.se/multimodal/papers/.

[8] T. Brondsted. Description of Visemes, March 1999, from http://www.kom.auc.dk/ tb/kurser/99gr872.htm.

[9] M. M. Cohen and D. W. Massaro. Modeling coarticulation in synthetic visual speech. In N. Thalmann and D Thalmann, editors, *Models and Techniques in Computer Animation*, pages 141–155. Springer-Verlag, Tokyo, 1993. Paper obtained from http://mambo.ucsc.edu/psl/ca93.html.

[10] *Don't Touch Me*, Short Animated Film. Kleiser-Walczak, Hollywood, CA, 1989.

[11] P. Ekman, editor. *Emotion in the Human Face, Second Edition.* Cambridge University Press, New York, NY, 1982.

[12] P. Ekman and W. V. Friesen. *Unmasking the Face.* Prentice-Hall, Englewood Cliffs, New Jersey, 1975.

[13] J. D. Foley, A. van Dam, S. K. Feiner, J. F. Hughes, and R. L. Phillips. *Introduction to Computer Graphics.* Addison-Wesley Publishing Company, Don Mills, Ontario, 1994.

[14] V. Hall. Synopsis of *Tin Toy*, November 1998, from http://mambo.ucsc.edu/psl/pixar.html.

[15] C. Kouadio, P. Poulin, and P. Lachapelle. Real-time facial animation based upon a bank of 3d facial expressions. In *Proceedings of Computer Animation 98*, pages 128–136, held in Philadelphia, Pennsylvania, June 1998. Paper obtained from http://www.iro.umontreal.ca/labs/infographie/papers/Kouadio-1998-RTFA/.

[16] J. Kulju, M. Sams, and K. Kaski. A finnish-talking head. In *Proceedings Finnic Phonetic Symposium*, August 1998. Paper obtained from http://www.lce.hut.fi/research/face/publications.html.

[17] Y. Lee, D. Terzopoulos, and K. Waters. Constructing physics-based facial models of individuals. In *Proceedings of Graphics Interface '93 Conference*, pages 1–8, Toronto, ON, May 1993. Pictures obtained from ftp://ftp.cs.toronto.edu/pub/dt/faceimg/.

[18] J. P. Lewis and F. I. Parke. Automated lip-synch and speech synthesis for character animation. In *Proceedings of Human Factors in Computing Systems and Graphics Interface*, Toronto, 1987.

[19] N. Magnenat-Thalmann, E. Primeau, and D. Thalmann. Abstract muscle action procedures for human face animation. *The Visual Computer*, 3(5):290–297, 1988. Picture obtained from http://ligwww.epfl.ch/˜thalmann/rdv.html.

[20] N. Magnenat-Thalmann and D. Thalmann. The direction of synthetic actors in the film *Rendez-vous à Montréal*. *IEEE Computer Graphics & Applications*, 7(12):9–19, 1987.

[21] K. Munhall. Personal Communication. August 1988. Web site: http://130.15.96.12/ munhallk/facial.html.

[22] F. I. Parke. Parameterized models for facial animation. *IEEE Computer Graphics and Applications*, 2(9):61–68, November 1982.

[23] F.I. Parke and K. Waters. *Computer Facial Animation.* A. K. Peters, Wellesley, MA, 1996. Code available as appendices at: http://www.crl.research.digital.com/publications/books/waters/waters_book.html.

[24] B. Rodriguez. Speech–a sight to behold. UCSC Science Notes (Summer 1996) from http://natsci.ucsc.edu/scicom/SciNotes/9601/Speech/00Intro.html.

[25] *Sextone for President*, Short Animated Film. Kleiser-Walczak, Hollywood, CA, 1988.

[26] D. Terzopoulos, B. Mones-Hattal, B. Hofer, F. Parke, D. Sweetland, and K. Waters. Facial animation: Past, present and future (panel summary). In *Proceedings of the ACM SIGGRAPH 97 Conference*, pages 434–436, Los Angeles, CA, August 1997.

[27] D. Terzopoulos and K. Waters. Analysis and synthesis of facial image sequences using physical and anatomical models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(6):569–579, June 1993. Picture modified from http://mambo.ucsc.edu/psl/fan.html.

[28] K. Waters. A muscle model for animating three-dimensional facial expressions. *Computer Graphics(SIGGRAPH'87)*, 21(4):17–24, July 1987.

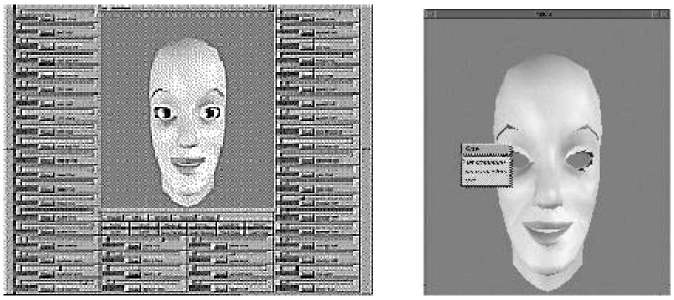
[29] K. Waters and T. M. Levergood. DECface: An automatic lip-synchronization algorithm for synthetic faces. Technical Report CRL 93/4, Digital Equipment Corporation Cambridge Research Lab, Cambridge, Massachusetts, September 1993. Obtained from: http://www.crl.research.digital.com-/publications/techreports/abstracts/93_4.html.

[30] J. R. Westbury. *X-Ray Microbeam Speech Production Database User's Handbook.* Waisman Center, University of Wisconsin, Madison, WI, June 1994.

[31] L. Williams. Performance-driven facial animation. *Computer Graphics*, 24(4):235–242, August 1990.

Appendix A

User Interface of CASSI



(a)with **(b)without**

Figure A.1: Marriot's Interface with and without the Forms Library

The user interface of CASSI is based on Marriot's code, which was obtained from "http://mambo.ucsc.edu-/psl/Fascia/". Marriot's code includes two interfaces: a forms library interface (shown in Figure A.1(a)) and a graphics library interface (shown in Figure A.1(b)). The forms library enables the creation of push buttons and sliding menu bars, which are shown surrounding the three side of Parke's face in Figure A.1(a). The sliding menu bars, in particular, enable the user to interactively adjust the parameters' values and view the corresponding changes in the face. The graphics library interface consists of a menu, which is shown as the square box on the left side of Figure A.1(b). This menu pops up when the user clicks on the right mouse button. Marriot's graphics library interface was inherited by CASSI. It was chosen over the forms library interface because it did not rely on the additional library (forms library), and because our application did not require manual adjustments on the parameter values.

To eliminate the code that relied on the forms library, we used Parke's code (obtained from "http:-//www.crl.research.digital.com/publications/books/waters/Appendix2/ap2.html") and augmented it with Marriot's graphics library interface and animation components. Besides the three menu options existing in Marriot's code, two additional options were added. The three options from Marriot's code were: "set orientation", which allows the face to be rotated from side to side, and up and down; "set parameters" (changed to "change/read parameters" in CASSI), which allows the parameters to be manually changed or read through standard input; and "quit" which exits the program. The two added options were: "animate", which allows the face to be animated according to one **.TXY* file that is input to standard input; and "display half/full", which allows half the face (cut vertically) to be displayed so that the inside of the mouth, particularly the tongue, can be viewed when the face is turned to the side. Figure A.2 shows the face and the CASSI menu with all five options.

Using the "set orientation" and "display half/full" options, the user can view the inside of the mouth, including the movement of the tongue. Figure A.3 shows a few of the views that can be obtained using these options. Figure A.3 (a) shows the face in the starting location without adjustments in its "orientation";

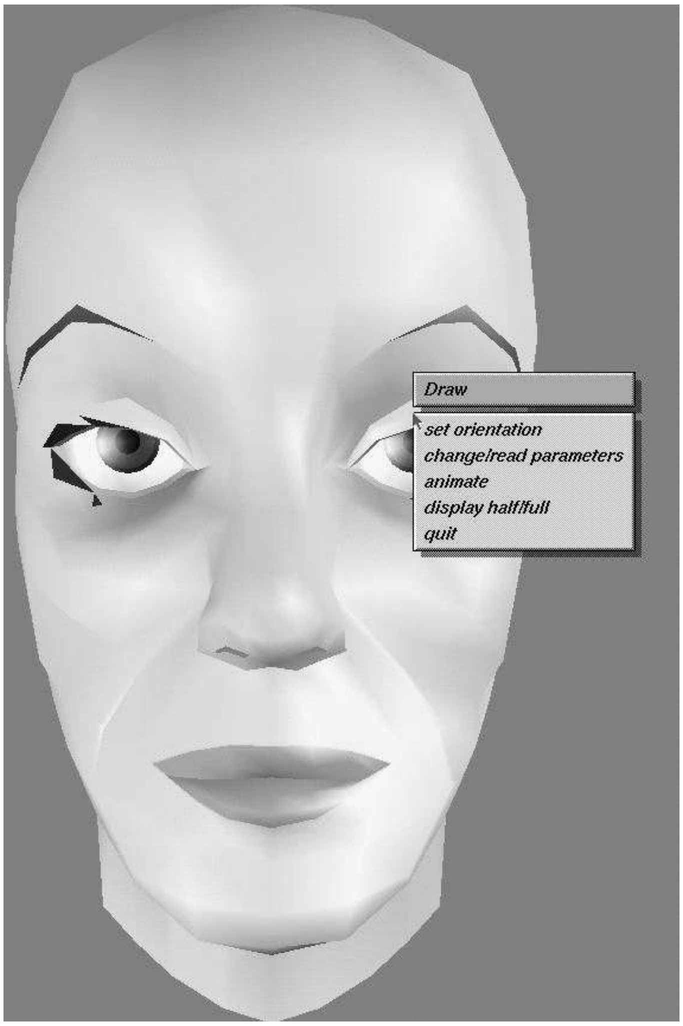
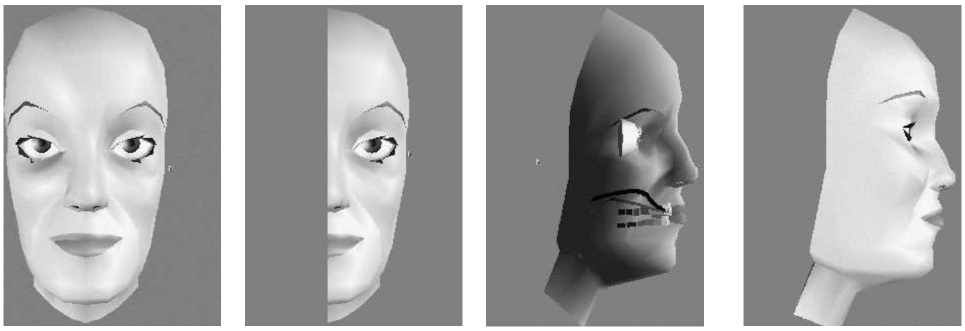


Figure A.2: The CASSI Interface

(b) shows the "cut in half" face, resulting from the "display half" option; (c) shows the rotated half face, resulting from the "set orientation" option and displaying the insides of the mouth, including the tongue, teeth, and palate; and (d) shows the restored "full" side of the face, resulting from the "display full" option. Three of these views ((a), (c), and (d)) are used in the subjective evaluation.



(a)Front View | (b)Half Face from Front | (c)Half Face from Side | (d)Side View

Figure A.3: Different Views of the Face

Appendix B

Files Used for Preprocessing XRMB data

To implement the lip rounding and lip thickness components of CASSI 2.0, which were described in Sections 4.2.1 and 4.2.3, preprocessing was required. This preprocessing extracted information from a sample of up to 20 **.txy* files for each subject. This appendix provides a list of the files used in this preprocessing. It also contains a description of the program used to extract information for the lip rounding component of CASSI 2.0.

Subject	Task Count	Task #'s
JW11	20	4,9,13,14,19,21,22,23,24,26,28,33,34,38,39,42,50,52,53,58
JW12	20	1,3,6,8,9,16,17,20,22,24,30,32,34,36,37,41,45,48,51,52
JW15	13	2,9,14,16,19,21,24,25,27,29,34,36,38
JW16	20	1,4,6,7,9,14,16,18,22,24,27,33,37,41,44,46,50,53,55,118
JW18	20	4,14,18,19,20,23,26,28,32,34,35,38,40,44,45,49,52,54,58,118_2
JW19	20	1,9,10,14,23,27,32,35,38,44,54,62,70,72,76,87,89,95,102,118
JW21	19	2,4,7,10,13,16,18,21,23,26,31,33,34,36,40,43,46,50,53
JW24	20	2,4,6,7,13,18,21,25,28,31,33,35,39,41,45,47,50,52,54,58
JW25	20	6,7,9,13,14,15,16,17,21,25,26,28,29,37,39,41,49,51,55,118
JW27	20	1,5,9,14,20,23,25,27,29,31,34,37,39,41,44,47,50,52,54,118
JW32	4	2,7,13,15
JW40	20	3,5,8,9,12,16,18,21,24,27,29,32,37,40,44,48,51,52,56,118
JW41	20	2,9,11,12,19,22,24,28,33,34,36,38,40,43,47,49,50,54,57,118
JW45	20	1,3,5,9,14,17,19,21,22,24,26,30,34,39,43,47,49,51,54,118
JW502	20	3,6,7,9,15,21,25,27,29,32,38,39,41,46,47,49,52,55,57,118

Table B.1: Subjects and Tasks used for Extracting Data on Lip Protrusion and Thickness

Table B.1 provides a summary of the up to 20 **.txy* files used for extracting information about lip rounding and lip thickness. The first column of the table provides the subject's name, the second column indicates the number of selected files, and the third column lists the task number for each of these files. In some cases, such as for JW32, fewer than 20 files are listed because 20 error-free files did not exist.

Information about the maximum, minimum, and starting x-positions was extracted from these files to eventually create the file *maxprotru.txt*, which was used for determining the key x values required for lip rounding. To extract the maximum, minimum, and starting x-positions, a program was created which scanned the up to 20 files for one subject and output relevant information. The algorithm of the program has the following steps: (1) for one file, save the first line of valid data as the starting location; (2) scan all subsequent lines of the file to determine the maximum and minimum x values of the upper and lower lip pellets; (3) repeat step 1 and 2 for the remaining selected **.txy* files corresponding to one subject; and (4) determine the average starting, average maximum, and average minimum x values over all the files. Figure B.1 shows the output of the program for subject JW11. Each row corresponds to one **.txy* file; thus,

```
JW11lipdata/
goodfile= 20
      STARTING        |       UPPER LIP         |       LOWER LIP         |
UL x        LL x      | Max X       Min X       | Max X       Min X       |
--------    --------    --------    --------    --------    --------
0.336931    0.227898    0.504508    0.309070    0.381411    0.105905
0.358329    0.259038    0.483463    0.306293    0.325273    0.114193
0.337849    0.226741    0.470944    0.307319    0.284569    0.072175
0.334403    0.236643    0.385501    0.317419    0.331342    0.095381
0.339454    0.235048    0.460202    0.316031    0.337587    0.044685
0.339207    0.238586    0.469578    0.322128    0.396667    0.069940
0.342600    0.214693    0.486451    0.302360    0.356205    0.083622
0.342443    0.230581    0.471346    0.010029    0.364903    0.003970
0.347668    0.223715    0.484564    0.295537    0.363921    0.075650
0.354284    0.234831    0.475649    0.323721    0.385286    0.080395
0.351460    0.241981    0.473498    0.325365    0.402808    0.119989
0.361439    0.233351    0.519758    0.293713    0.370653    0.061772
0.342133    0.206028    0.456724    0.319531    0.370406    0.093276
0.344332    0.221365    0.433136    0.324204    0.392955    0.077897
0.350138    0.218098    0.469897    0.009925    0.388293    0.003348
0.342745    0.215112    0.458812    0.301870    0.356276    0.099246
0.347092    0.206861    0.496364    0.309022    0.365978    0.076221
0.344968    0.240663    0.493773    0.309182    0.375274    0.060933
0.351081    0.236348    0.439413    0.319578    0.370886    0.143783
0.336922    0.226213    0.434701    0.312058    0.439836    0.125085
---------------------------averages----------------------------------
0.345274    0.228690    0.468414    0.281718    0.368026    0.080373
```

Figure B.1: Starting, Maximum, and Minimum X Positions for Subject JW11

JW11's summary has 20 rows corresponding to 20 error-free **.txy* files. The final row contains the averages for each of the columns. The first and second columns consist of the starting x values of the upper and lower lip respectively; these values are the result of step 1 in the algorithm. The other four columns contain the results of step 2, which scans one file for the maximum and minimum x-positions, where the results for the upper lip are stored in the third and fourth columns and and those for the lower lip are in the fifth and sixth columns. All values in Figure B.1 have been normalized, or divided by the XRMB jaw length, to eliminate potential scaling differences. Summaries similar to the one in Figure B.1 were obtained for each of the 15 subjects.